

# IBM Research Report

## Multidimensional Balanced Allocation for Multiple Choice & $(1 + \beta)$ Processes

**Ankur Narang**

IBM Research Division  
IBM India Research Lab  
Plot - 4, Block - C, ISID Campus, Vasant Kunj  
New Delhi - 110070, India.  
annarang@in.ibm.com

**Sourav Dutta**

IBM Research Division  
IBM India Research Lab  
Plot - 4, Block - C, ISID Campus, Vasant Kunj  
New Delhi - 110070, India.  
sodutta3@in.ibm.com

**Souvik Bhattacharjee**

IBM Research Division  
IBM India Research Lab  
Plot - 4, Block - C, ISID Campus, Vasant Kunj  
New Delhi - 110070, India.  
souvikbh@in.ibm.com

**IBM Research Division**

**Almaden - Austin - Beijing - Delhi - Haifa - T.J. Watson - Tokyo - Zurich**

**LIMITED DISTRIBUTION NOTICE:** This report has been submitted for publication outside of IBM and will probably be copyrighted is accepted for publication. It has been issued as a Research Report for the early dissemination of its contents. In view of the transfer of copyright to the outside publisher, its distribution outside of IBM prior to publication should be limited to peer communications and specific requests. After outside publication, requests should be filled only by reprints or legally obtained copies of the article (e.g., payment of royalties). Copies may be requested from IBM T.J. Watson Research Center, Publications, P.O. Box 218, Yorktown Heights, NY 10589 USA (email: reports@us.ibm.com). Some reports are available on the internet at <http://domino.watson.ibm.com/library/CyberDig.nsf/home>.

arXiv:1111.0762v2 [cs.DS] 29 Nov 2011

## Abstract

Allocation of balls into bins is a well studied abstraction for load balancing problems. The literature hosts numerous results for sequential (single dimensional) allocation case when  $m$  balls are thrown into  $n$  bins; such as: for multiple choice paradigm the expected gap between the heaviest bin and the average load is  $O(\frac{\log \log(n)}{\log(d)})$  [4],  $(1 + \beta)$  choice paradigm with  $O(\frac{\log(n)}{\beta})$  gap [10] as well as for single choice paradigm having  $O(\sqrt{\frac{m \log(n)}{n}})$  gap [9]. However, for multidimensional balanced allocations very little is known. Mitzenmacher [6] proved  $O(\log \log(nD))$  gap for the multiple choice strategy and  $O(\log(nD))$  gap for single choice paradigm (where  $D$  is the total number of dimensions with each ball having exactly  $f$  populated dimensions) under the assumption that for each ball  $f$  dimensions are uniformly distributed over the  $D$  dimensions. In this paper we study the symmetric multiple choice process for both unweighted and weighted balls as well as for both multidimensional and scalar modes. Additionally, we present the results on bounds on gap for the  $(1 + \beta)$  choice process with multidimensional balls and bins.

In the first part of this paper, we study multidimensional balanced allocations for the symmetric  $d$  choice process with  $m \gg n$  unweighted balls and  $n$  bins. We show that for the symmetric  $d$  choice process and with  $m = O(n)$ , the upper bound (assuming uniform distribution of  $f$  populated dimensions over  $D$  total dimensions) on the gap is  $O(\ln \ln(n))$  w.h.p.. This upper bound on the gap is within  $D/f$  factor of the lower bound. This is the first such tight result along with detailed analysis for  $d$  choice paradigm with multidimensional balls and bins. This improves upon the best known prior bound of  $O(\log \log(nD))$  [6]. For the general case of  $m \gg n$  the expected gap is bounded by  $O(\ln \ln(n))$ . For variable  $f$  and non-uniform distribution of the populated dimensions (using analysis for weighted balls), we obtain the upper bound on the expected gap as  $O(\log(n))$ .

Further, for the multiple round parallel balls and bins, using symmetric  $d$ -choice process in multidimensional mode, we show that the gap is also bounded by  $O(\log \log(n))$  for  $m = O(n)$ . The same bound holds for the expected gap when  $m \gg n$ .

Our analysis also has the following strong implications for the sequential scalar case. For the weighted balls and bins and general case  $m \gg n$ , we show that the upper bound on the expected gap is  $O(\log(n))$  (assuming  $E[W] = 1$  and second moment of the weight distribution is finite) which improves upon the best prior bound of  $n^c$  ( $c$  depends on the weight distribution that has finite fourth moment) provided in [12]. Our analysis also provides a much easier and elegant proof technique (as compared to [4]) for the  $O(\log \log(n))$  upper bound on the gap for scalar unweighted  $m \gg n$  balls thrown into  $n$  bins using the symmetric multiple choice process.

Moreover, we study multidimensional balanced allocations for the  $(1 + \beta)$  choice process and the multiple ( $d$ ) choice process. We show that for the  $(1 + \beta)$  choice process and  $m = O(n)$  the upper bound (assuming uniform distribution of  $f$  populated dimensions over  $D$  total dimensions) on the gap is  $O(\frac{\log(n)}{\beta})$ , which is within  $D/f$  factor of the lower bound. For fixed  $f$  with non-uniform distribution and for random  $f$  with Binomial distribution the expected gap remains  $O(\frac{\log(n)}{\beta})$  and is independent of the total number of balls thrown,  $m$ . This is the first such tight result along with detailed analysis for  $(1 + \beta)$  paradigm with multidimensional balls and bins.

## 1 Introduction

Balls-into-bins processes serve as a useful abstraction for resource balancing tasks in distributed and parallel systems. Assume  $m$  balls are to be put sequentially into  $n$  bins, where typically the goal is to minimize the load, measured by the number of balls, in the most loaded bin. In the classic *single choice* process each ball is placed in a bin chosen independently and uniformly at random. For the case of  $n$  bins and  $m = n$  balls it is well known that the load of the heaviest bin is at most  $(1 + o(1)) \frac{\ln(n)}{\ln \ln(n)}$  balls with high probability (w.h.p.). Further, if  $m > n \ln(n)$  then the load in the heaviest bin is given by at most  $\frac{m}{n} + \sqrt{\frac{m \log(n)}{n}}$  [11].

An interesting and substantial decrease in the maximum load is achieved by the use of the *multiple choice* paradigm (also referred to as *d choice* paradigm), given as: Let  $Greedy(U, d)$  denote the algorithm where each ball is inserted into the lesser loaded among the  $d \geq 2$  bins, independently sampled from  $U$ , where  $U$  denotes the uniform distribution over the bins. In a seminal paper Azar et. al. [3] proved that when  $m = n$  and the balls are inserted by  $Greedy(U, d)$  the heaviest bin has load of  $\frac{\ln \ln(n)}{\ln(d)} + \theta(1)$  w.h.p.. The case for  $d = 2$  was proved by Karp et.al. in [8], later being generalized by Berenbink et.al. [4] to prove the following:

**Theorem 1.1** *Let  $\gamma$  denote a suitable constant. If  $m$  balls are allocated into  $n$  bins using  $Greedy(U, d)$  with  $d \geq 2$  then the number of bins with load at least  $\frac{m}{n} + i + \gamma$  is at most  $n \cdot \exp(-d^i)$  with probability at least  $1 - 1/n$ . ([4])*

An immediate corollary is that w.h.p. the heaviest bin has a load of  $\frac{m}{n} + \frac{\log \log(n)}{\log(d)} + O(1)$ . Thus, the additive gap between the maximum load and the average load is *independent* of the number of balls thrown.

The multiple-choice paradigm and balls-and-bins models have several interesting applications. In particular, the two-choice paradigm can be used to reduce the maximum time required to search a hash table. If instead of using

a single perfectly random hash function as in a typical hash table implementation (with maximum chain length as  $O(\ln(n))$ ), we use two perfectly random hash functions, then the length of the longest chain reduces to  $O(\ln \ln(n))$ . In efficient PRAM simulation, the two-choice paradigm helps in reducing the contention ([2]) of processors to access the same memory (DRAM). Further, the two-choice scheme can be advantageous in situations, for example when one hopes to fit one full chain in a single cache line [5]. The multiple-choice approach has also proven useful in online (dynamic) assignment of tasks to servers (disk servers, network servers etc). By using multiple-choice one would get much better load balance across the servers as compared to the single-choice approach.

In many practical problems, the underlying data can be *multidimensional*. This is especially true for parallel data mining and machine learning problems, where the input data has many dimensions such as text search where the distinct words in the document set can be considered as the dimensions and the total number of dimensions equals the size of the vocabulary that could potentially run into millions of words. Because the collection of pages to be indexed is so large, it has to be split among  $n$  servers. When a user makes a query to a front-end machine, the query is sent to all  $n$  servers; results are returned to the front-end machine for merging and presentation. Hence, the time to serve the query is determined by the slowest of the servers, the critical process. The time for each server to process a one-word query is roughly proportional to the number of documents at that server containing the word of interest. Thus, to achieve better efficiency, it is necessary to efficiently split the documents among servers in such way that the number of documents containing a given word is roughly equal.

Further, many application domains such as Telecommunication, Finance and others also involve huge number of dimensions such as genres and sub genres of songs and videos for collaborative filtering<sup>1</sup> type correlational analysis between the users. Here, one would like to predict what type of item (song or video) one user could prefer based on his inferred relationship with other similar users. Due to *massive size* of the multidimensional data in such distributed data mining and machine learning problems, one needs to devise *online load balancing* algorithms. While, dimensionality reduction techniques can reduce the total number of dimensions to work on, even then, one needs to handle data with large number of dimensions. Further, this data is highly sparse, i.e. number of filled entries in the (*user \* item*) matrix is a small fraction of total possible entries in the matrix. Thus, distributed data mining and machine learning (for example in cloud computing environments), suffer from severe scalability and parallel efficiency issues due to huge load imbalance across the machines in the compute cluster (cloud). Hence, there is a strong need to address load balancing for multidimensional datasets.

## 1.1 Probability Distribution for Bin Selection

The  $d$ -choice scheme can be characterized by a probability vector  $p = (p_1, p_2, p_3, \dots, p_n)$ , where  $p_i$  denotes the probability a ball falls in the  $i^{th}$  most loaded bin. Here, the bins are ordered from the most loaded to the least loaded (ties are broken arbitrarily). Then,  $p_1$  denotes the probability that the most loaded bin receives the current ball,  $p_2$  denotes the probability that the second bin (in the order) receives the ball and so on. In general, in the  $d$ -choice scheme,  $p_i = (\frac{i}{n})^d - (\frac{i-1}{n})^d$ . For  $d = 1$ ,  $\forall i : p_i = 1/n$ , while for  $d > 1$ ,  $p_i > p_j$  for  $i > j$ . Thus, for  $d > 1$ , the process has bias towards the lighter bins. This biasing leads to an overall lower gap for  $d > 1$  choice as compared to single choice ( $d = 1$ ) process.

In this paper, we consider the multidimensional variant of the balls and bins problem. One multidimensional variant, proposed by [6] is as follows: Consider throwing  $m$  balls into  $n$  bins, where each ball is a uniform  $D$ -dimensional 0-1 vector of weight  $f$ . Here, each ball has exactly  $f$  non-zero entries chosen uniformly among all  $\binom{D}{f}$  possibilities (Fig. 1 (in the Appendix A)). The average load in each dimension for each bin is given as  $mf/nD$ . Let  $l(a, b)$  be the load in the dimension  $a$  for the  $b^{th}$  bin. The gap in a dimension (across the bins) is given by  $gap(a) = \max_b l(a, b) - avg(a)$ , where  $avg(a)$  is the average load in the dimension  $a$ . The maximum gap across all the dimensions,  $\max_a gap(a)$ , then determines the load balance across all the bins and the dimensions. Thus, for the multidimensional balanced allocation problem, the objective is to minimize the maximum gap (across any dimension). We refer to the multidimensional ball as *md-ball* and the multidimensional bin as *md-bin*.

In another variation of multidimensional balanced allocation the constraint of uniform distribution for populated entries is removed. Here again, each ball is a  $D$  dimensional 0-1 vector and each ball has exactly  $f$  populated dimensions, but these populated dimensions can have an arbitrary distribution. In the third variation that is most general of the three, the number of populated dimensions,  $f$ , may be different across the balls, where  $f$  then is a random variable with an appropriate distribution.

Mitzenmacher et.al. in [6] addressed both the single choice and  $d$ -choice paradigm for multidimensional balls and bins under the assumption that balls are uniform  $D$ -dimensional  $(0, 1)$  vectors, where each ball has exactly  $f$  populated dimensions. They show that the gap for multidimensional balls and bins, using the two-choice process, is bounded

<sup>1</sup>[http://en.wikipedia.org/wiki/Collaborative\\_filtering](http://en.wikipedia.org/wiki/Collaborative_filtering)

by  $O(\log \log(nD))$ . However, this result is not tight and assumes that  $f$  is  $\text{polylog}(n)$ . Due to arbitrary number of dimensions and the resulting discrepancy across the dimensions along with the general case of  $m \gg n$ , the balanced allocations for multidimensional balls and bins is a challenging problem. In this paper, we compute bounds on the gap for the symmetric  $d$ -choice process for multidimensional balls and bins for both sequential and parallel scenarios.

## 1.2 Summary of Key Results & Techniques

We present detailed analysis for the online sequential and parallel multidimensional balls and bins using the symmetric  $d$ -choice process and show that for  $n$  bins and  $m = O(n)$  balls, the gap (assuming that exactly  $f$  populated dimensions are uniformly distributed over  $D$  per ball) achieved is  $O(\ln \ln(n))$ . We establish the first ever known bound for  $d$ -choice process and also show that this bound is tight (within  $D/f$  factor) by providing the lower bound as well. This improves upon the best prior bound of  $O(\log \log(nD))$  [6]. For the general case of  $m \gg n$ , the upper bound on the gap is  $O((\frac{mf}{nD})^{(1/2+\zeta)} \ln \ln(n))$  w.h.p., while the expected gap is still  $O(\ln \ln(n))$ . For non-uniform distribution with fixed  $f$  and for variable  $f$  with binomial distribution, we show that the expected gap is still independent of  $m$ .

In order to arrive at these results, a novel generic potential function based approach along with *sum load* across the dimensions per bin is used. This is much more challenging than the analysis presented by [10] for  $(1 + \beta)$ -choice process, as we obtain a much tighter bound of  $O(\ln \ln(n))$  (as compared to  $O(\frac{\log(n)}{\beta})$  in [10]). This requires a novel potential function as well as a much tighter analysis in each lemma to ensure that the expected value of the potential function is less than  $O(\ln(n))$  at all time  $t$  and satisfies the super-martingale property.

For parallel multidimensional balls and bins with multiple rounds using the  $d$  choice process, we show the upper bound on the gap as  $O(\log \log(n))$  for  $m = O(n)$  balls; and extend this bound on the gap to the general case of  $m \gg n$ . This is tighter than the  $O(\log \log(nD))$  that can be obtained using the analysis similar to [6].

For the weighted and heavy case ( $m \gg n$ ) using symmetric multiple choice sequential process for scalar balls, we prove an upper bound of  $O(W^* \log(n))$  (where  $W^*$  is the expected weight of the distribution), which improves upon the best prior bound of  $O(n^c)$  [12]. Our analysis technique also provides an alternate proof for the symmetric  $d$ -choice process with scalar unweighted  $m \gg n$  balls into  $n$  bins, that is simpler and elegant as compared to [4].

Further, we present the analysis for bounds on the gap for  $(1 + \beta)$  choice multidimensional process and prove that for  $m = O(n)$  the upper bound on the gap is  $O(\frac{\log(n)}{\beta})$  w.h.p. for uniform distribution of  $f$  dimensions over the  $D$  dimensions. For non-uniform distribution with fixed  $f$  and also for variable  $f$  the expected gap is  $O(\frac{\log(n)}{\beta})$  which is independent of  $m$ . Table 1 summarizes the comparison between our upper bounds and the best known prior bounds, with key results highlighted.

Process: $d$ -choice	Best Prior Bound	Our Bound
Multidim, Fixed- $f, m = O(n)$	$O(\log \log(nD))$ [6]	$O(\ln \ln(n))$
<b>Multidim, Fixed-<math>f, m \gg n</math></b>	None	$O(\ln \ln(n))$ (expected)
<b>Multidim, Var-<math>f</math></b>	None	$O(\log(n))$ (expected)
<b>Weighted Scalar, <math>m \gg n</math></b>	$O(n^c)$ (short memory via coupling [12])	$O(\log(n))$
Unweighted Scalar, $m \gg n$	$O(\ln \ln(n))$ (using layered induction and short memory [4])	$O(\ln \ln(n))$ (using potential function)
Parallel Multidim, $m = O(n)$	$O(\log \log(nD))$ (adaptation of [6])	$O(\ln \ln(n))$
<b>Parallel Multidim, <math>m \gg n</math></b>	None	$O(\ln \ln(n))$ (expected)
Parallel Scalar	$O(\ln \ln(n))$ (for $m = O(n)$ [1])	$O(\ln \ln(n))$ (for $m \gg n$ )
Process: $(1 + \beta)$ -choice	Best Prior Bound	Our Bound
Multidim, Fixed- $f, m = O(n)$	None	$O(\frac{\log(n)}{\beta})$
<b>Multidim, Fixed-<math>f, m \gg n</math></b>	None	$O(\frac{\log(n)}{\beta})$ (expected)
<b>Multidim, Var-<math>f</math></b>	None	$O(\frac{\log(n)}{\beta})$ (expected)

Table 1: Upper Bound Comparison for  $d$ -choice and  $(1 + \beta)$  Process

## 2 Related Work

Balls into bins is a well studied abstraction for load balancing problems. Numerous results are known for sequential (single dimensional) allocation case when  $m$  balls are thrown into  $n$  bins; such as: for multiple choice paradigm

the expected gap between the heaviest bin and the average load is  $O(\frac{\log \log(n)}{\log(d)})$  [4],  $(1 + \beta)$  choice paradigm with  $O(\frac{\log(n)}{\beta})$  gap [10] as well as for single choice paradigm having  $O(\sqrt{\frac{m \log(n)}{n}})$  gap [9]. [3] showed that the bound of  $O(\frac{\log \log(n)}{\log(d)})$  for the symmetric  $d$  choice process is stochastically optimal, i.e. any other greedy approach using the placement information of the previous balls to place the current ball majorizes to their approach. However, if the alternatives are drawn from different groups then different rules for tie breaking result in different allocations. [13] presents such an *asymmetric* strategy and using witness tree based analysis proves that this leads to improvement in load balance to  $O(\frac{\log \log(n)}{d \log(\phi_d)})$  w.h.p. where,  $\phi_2$  is the golden ratio and  $\phi_d$  is a simple generalization.

The multiple choice and in particular the two-choice paradigm and balls-and-bins models have several interesting applications. In particular, the two-choice paradigm can be used to reduce the maximum search time in a hash table. Instead of using a single perfectly random hash function as in a typical hash table implementation (with maximum chain length as  $O(\ln(n))$ ), if we use two perfectly random hash functions, then the length of the longest chain reduces to  $O(\ln \ln(n))$ . In the latter case, when inserting a key, we apply both hash functions to determine the two possible table entries where the key can be inserted. Then, of the two possible entries, we add the key to the shorter of the two chains. To search for an element, we have to search through the chains at the two entries given by both hash functions. If  $n$  keys are sequentially inserted into the table, the length of the longest chain is  $O(\log \log n)$  with high probability, implying that the maximum time needed to search the hash table is  $O(\log \log n)$  with high probability. Further, the two-choice scheme can be advantageous in situations for example, when one hopes to fit one full chain in a single cache line [5]. The two-choice approach has also proven useful in online (dynamic) assignment of tasks to servers (disk servers, network servers etc). By using two-choice one would get much better load balance across the servers as compared to the single-choice approach. If we use  $(1 + \beta)$  choice then, we would get around  $\log(n)$  gap (as compared to  $O(\log \log(n))$  gap for the two-choice) but the communication cost to query the load of the servers will be lesser by  $(1 + \beta)/2$  factor as compared to the two-choice approach.

Cole et al. [7] show that the two-choice paradigm can be applied effectively in a different context, namely, that of routing virtual circuits in interconnection networks with low congestion. They show how to incorporate the two-choice approach to a well-studied paradigm due to Valiant for routing virtual circuits to achieve significantly lower congestion.

Kunal et.al. [10] present that for online sequential  $(1 + \beta)$  choice process with  $n$  bins and  $m \gg n$  balls, a tight gap of  $O(\frac{\log(n)}{\beta})$  can be obtained. They use a potential function based technique and further use a majorization argument to generalize their result. We present a novel generic potential function based approach with *sum load* function across all dimensions of a bin for multidimensional balls and bins and obtain tight bounds on the gap for the  $d$ -choice process for both sequential and parallel scenarios. Our analysis is much more challenging than [10] since we prove a tighter bound that requires a much tighter analysis in each lemma to prove that the expected value of potential function is less than  $O(\ln(n))$  at all time  $t$ . Further, the lower and upper bounds for the  $(1 + \beta)$  choice process with multidimensional balls and bins have also been provided in this paper.

Mitzenmacher et.al. in [6] address both the single choice and  $d$ -choice paradigm for multidimensional balls and bins under the assumption that balls are uniform  $D$ -dimensional  $(0, 1)$  vectors, where each ball has exactly  $f$  populated dimensions. They show that the gap for multidimensional balls and bins, using the two-choice process, is bounded by  $O(\log \log(nD))$ . We provide better bound on the gap ( $O(\log \log(n))$ ) and also provide the bound for the general case of  $m \gg n$ . Further, while [6] assumes that  $f$  is *polylog*( $n$ ) we don't make any such assumptions. For the multiple round multidimensional parallel balls and bins process where in each round, each bin accepts at max only a single ball, one can use layered induction based proof [6] to get a similar bound on the gap as  $O(\log \log(nD))$ . Using our novel potential function based analysis we show a tighter upper bound of  $O(\log \log(n))$ . The bound for the general case of  $m \gg n$  is also provided.

Berenbrink et.al. [4] prove an upper bound of  $O(\log \log(n))$  for the general case of  $m \gg n$  balls and  $n$  bins using a sophisticated analysis involving two main steps. In the first step, they show that when the number of balls is polynomially bounded by the number of bins the gap can be bounded by  $O(\ln \ln(n))$ , using the concept of layered induction and some additional tricks. In particular, they consider the entire distribution of the bins in the analysis (while in typical  $m = O(n)$  case the bins with load smaller than the average could be ignored). In the second step, they extend this result to general  $m \gg n$  case, by showing that the multiple-choice processes are fundamentally different from the classical single-choice process in that they have *short memory*. This property states that given some initial configuration with gap  $\Delta$ , after adding *poly*( $n$ ) more balls the initial configuration is *forgotten*. The proof of the short memory property is done by analyzing the mixing time of the underlying Markov chain describing the load distribution of the bins. The study of the mixing time is via a new variant of the coupling method (called *neighboring coupling*). We prove the same result on the gap ( $O(\log \log(n))$ ) for the symmetric  $d$  choice process with  $m \gg n$  but by using a much simpler and elegant potential function based approach.

Kunal et.al. [12] prove that for weighted balls (weight distribution with finite fourth moment) and  $m \gg n$ , the expected gap is independent of the number of balls and is less than  $n^c$ , where  $c$  depends on the weight distribution. They first prove the weak gap theorem which says that w.h.p  $\text{Gap}(t) < t^{2/3}$ . Since in the weighted case the  $d$  choice process is not dominated by the one choice process, they prove the weak gap theorem via a potential function argument. Then, the *short memory theorem* is proved. While in [4] the short memory theorem is proven via coupling, [12] uses similar coupling arguments but defines a different distance function and use a sophisticated argument to show that the coupling converges. [12] also presents a reduction from the real-weighted case to the integer-weighted case. We present the results for weighted case (with integer and real weights and weight distribution with finite second moment) using an elegant and much simpler potential function based argument and show that the gap for arbitrary  $m \gg n$  is bounded by  $O(W^* \log(n))$ , where  $W^*$  is the expected weight of the distribution. Adler et.al. [1] consider parallel balls and bins with multiple rounds. They present analysis for  $O(\frac{\log \log(n)}{\log(d)})$  bound on the gap (for  $m = O(n)$ ) using  $O(\frac{\log \log(n)}{\log(d)} + O(d))$  rounds of communication. We generalize this result to the case of parallel multidimensional balls and bins and arbitrary  $m \gg n$  balls with similar bound on the gap.

### 3 Symmetric $d$ -choice Process

In this section, we present various results on the bounds on the gap using the symmetric  $d$ -choice process including unweighted sequential and parallel multidimensional balls and bins and the sequential weighted scalar case.

#### 3.1 Markov Chain Specification

As mentioned earlier, a balls-and-bins process can be characterized by a probability distribution vector  $(p_1, p_2, p_3, \dots, p_n)$ , where,  $p_i$  is the probability a ball is placed in the  $i^{\text{th}}$  most loaded multidimensional bin. Let  $x_i^d(t)$  be the random variable, that denotes the *weight in dimension  $d$  for bin  $i$*  and is equal to the load of the  $d^{\text{th}}$  dimension of the  $i^{\text{th}}$  bin minus the average load in dimension  $d$ . So,  $\sum_{i=1}^n x_i^d(t) = 0, \forall d \in [1..D]$ . Each md-ball has  $f$  populated dimensions, where  $f$  could be constant across the balls or a random variable with a given distribution. Let,  $s_i(t)$  denote the sum of the loads (minus corresponding dimension averages) across all  $D$  dimensions for the bin  $i$  at time  $t$ , expressed as  $s_i(t) = \sum_{d=1}^D x_i^d(t)$ . It is assumed that bins are sorted by  $s_i(t)$ . So,  $s_i \geq s_{i+1} \forall i \in [1..n-1]$ . The process defines a Markov chain over the matrices,  $x(t)$  as follows:

- Sample  $j \in_p [n]$ .
- Set  $r_i = s_i(t) + f(1 - 1/n)$ , for  $i = j$ . Since, an md-ball has  $f$  non-zero entries, so each of these  $f$  dimensions in the bin,  $i$ , will be incremented by  $1 - 1/n$ .
- Set  $r_i = s_i(t) - f/n$ , for  $i \neq j$ . Since, an md-ball has  $f$  non-zero entries, so the each of the corresponding  $f$  dimensions in the bin,  $i$ , will be decremented by  $1/n$ . This ensures that for each dimension the sum across all the bins is 0.
- Obtain  $s(t+1)$  by sorting  $r(t)$ .

Fig. 1 (in the Appendix A) illustrates a multidimensional balls and bins scenario. The bounds on the gap will be proven for a family of probability distribution vectors  $p$ . As mentioned earlier, the md-bins are sorted based on their total dimensional load, i.e. sum of the weights across all dimensions for each bin ( $s_i$  for bin  $i$ ).

In the remaining analysis, we assume that when an md-ball arrives, then the selection of the bins is based on  $s_i$ , i.e. total sum of weights across all dimensions for the randomly selected bins (Fig. 1 in Appendix A). In particular, for the  $d = 2$  choice process, when  $d$  bins are randomly selected, the md-ball (with  $f$  non-zero entries) is assigned to the md-bin with the lowest  $s_i$ . Using this selection mechanism, we prove the upper and lower bounds on the gap obtained for the  $d$  choice process. Note that, this is a different allocation mechanism than that considered in [6] where the *max* criteria is used over the restricted set of  $f$  populated dimensions in the current md-ball. Further, we prove upper bound for the case when  $m \gg n$ , while [6] considers  $m = O(n)$  case. The proofs below hold for even the case when  $d > 2$ , though we consider the case for  $d = 2$  for sake of clarity.

#### 3.2 Upper Bound On Gap for Unweighted Case

Let there be some constants,  $\epsilon > 0, \theta > 1, \gamma_1$  and  $\gamma_2, \gamma_3, \gamma_4$ , where,  $0 < \gamma_1 < \gamma_3 < 1/2 < \gamma_4 < \gamma_2 < 1$  and  $\theta\gamma_1 = 1; \gamma_1 + \gamma_2 = 1$ , and  $\gamma_3 + \gamma_4 = 1$ . Since we consider the  $d$ -choice process, the probability of selecting the bins has strong

bias in favor of the lightly loaded bins. For  $d = 2$ , this results in the following:

$$\begin{aligned} p(n\gamma_3) &\leq \frac{2n\gamma_3 - 1}{n^2} \\ p(n\gamma_4) &\geq \frac{2n\gamma_4 - 1}{n^2} \end{aligned}$$

This implies that:  $\sum_{i \geq (n\gamma_2)} p_i \geq (1 - \gamma_2^2)$  and  $\sum_{i \leq (n\gamma_1)} p_i \leq \gamma_1^2$ . We assume that  $\epsilon \leq 1/4$ . Further, let  $\alpha = \epsilon/2f$ . In the analysis below, we assume each md-ball has exactly  $f$  populated dimensions ( $f$  constant, fixed- $f$  case). This is similar to the *unweighted case* with scalar balls.

The md-bins can be arranged in a partial order, according to their  $s_i$  values. Define an *equi-load* group (say  $p$ ) as a set of bins with the same  $s_i$  value. Define the potential of an equi-load group ( $p$ ) as  $\Phi(G_p) = \sum_{k=0}^{|G_p|-1} \frac{e^{\alpha \cdot s_k}}{p_0 + k}$ , where  $p_0$  is the beginning index for the group,  $|G_p|$  is the size of the  $p^{th}$  group and  $e^{\alpha \cdot s_k} = e^{\alpha \cdot s_{k+1}}, \forall k \in [p_0..(p_0 + |G_p| - 2)]$ . The  $n$  bins are partitioned into disjoint equi-load groups (total  $|G|$  groups), i.e. each bin is assigned to only a single equi-load group. The group structure defined here helps in characterizing the change in index of the bin that gets the ball (after sorting).

Similarly, define another potential function for an equi-load group as,  $\Psi(G_p) = \sum_{k=0}^{|G_p|-1} \frac{e^{-\alpha \cdot s_k}}{p_0 + k}$ . Now, define the following potential functions over all the groups:

$$\begin{aligned} \Phi(t) &= \Phi(s(t)) = \sum_{p=1}^{|G|} \Phi(G_p) \\ \Psi(t) &= \Psi(s(t)) = \sum_{p=1}^{|G|} \Psi(G_p) \\ \Gamma(t) &= \Gamma(s(t)) = \Phi(t) + \Psi(t) \\ &= \sum_{i=1}^n \left[ \frac{e^{\alpha \cdot s_i}}{i} + \frac{e^{-\alpha \cdot s_i}}{i} \right] \end{aligned} \tag{3.1}$$

where,  $s_i(t) = \sum_{d=1}^D x_i^d(t)$

In the beginning, each dimension for each bin has 0 weight, thus  $s_i = 0, \forall i$  and hence,  $\Gamma(0) = 2 \ln(n)$ . We show that if  $\Gamma(x(t)) \geq a \ln(n)$  for some  $a > 0$ , then  $\mathbb{E}[\Gamma(t+1)|x(t)] \leq (1 - \frac{\epsilon}{8n(1+\epsilon\gamma_1)}) * \Gamma(t)$ . This helps in demonstrating that for every given  $t$ ,  $\mathbb{E}[\Gamma(t)] \in O(\ln(n))$ . This implies that the maximum gap is  $O(\ln \ln(n))$  w.h.p.

First, consider the change in  $\Phi(t)$  (also refers to  $\Phi$  by default) and  $\Psi(t)$  (also refers to  $\Psi$  by default) separately when a ball is thrown with the given probability distribution.

**Lemma 3.1** When an md-ball is thrown into an md-bin, the following inequality holds:

$$\mathbb{E}[\Phi(t+1) - \Phi(t)|x(t)] \leq \sum_{i=1}^n [p_i * (\alpha \cdot f + (\alpha \cdot f)^2) - \alpha \cdot f/n] \cdot e^{\alpha \cdot s_i} \tag{3.2}$$

**Proof** Let  $\Delta_i$  be the expected change in  $\Phi$  if the ball is put in bin,  $i$ . So,  $r_i(t+1) = s_i + f(1 - 1/n)$ ; and for  $j \neq i$ ,  $r_j(t+1) = s_j(t) - f/n$ . The new values i.e.  $s(t+1)$  are obtained by sorting  $r(t+1)$  and  $\Phi(s) = \Phi(r)$ . When, an md- ball is committed to bin  $i$ , then it moves to the end of the previous equi-load group or it creates a new equi-load group and hence can be located at index  $p_0$  (beginning location of its prior group) in the new sorted order of the bins. Thus, the expected contribution of bin,  $i$ , to  $\Delta_i$  is given as follows:

$$\mathbb{E}\left[\frac{e^{\alpha \cdot (s_i + f(1-1/n))}}{p_0}\right] - \frac{e^{\alpha \cdot s_i}}{p_0} = \frac{e^{\alpha \cdot s_i}}{p_0} [e^{\alpha \cdot f(1-1/n)} - 1]$$

Similarly, the expected contribution of bin,  $j$  ( $j \neq i$ ) to  $\Delta_i$  is given as:

$$\mathbb{E}\left[\frac{e^{\alpha \cdot (s_j - f/n)}}{j}\right] - \frac{e^{\alpha \cdot s_j}}{j} = \frac{e^{\alpha \cdot s_j}}{j} [e^{-\alpha \cdot f/n} - 1]$$

Therefore,  $\Delta_i$  is given as follows:

$$\begin{aligned} \Delta_i &= \Phi_i[e^{\alpha \cdot f(1-1/n)} - 1] + \sum_{j \neq i} \Phi_j[e^{-\alpha \cdot f/n} - 1] \\ &= \Phi_i e^{-\alpha \cdot f/n} (e^{\alpha \cdot f} - 1) + (e^{-\alpha \cdot f/n} - 1) \cdot \Phi \end{aligned}$$

Thus, we get the overall expected change in  $\Phi$  as follows:

$$\begin{aligned}
\mathbb{E}[\Phi(t+1) - \Phi(t)|x(t)] &= \sum_{i=1}^n p_i * \Delta_i \\
&= \sum_{i=1}^n p_i * [\Phi_i e^{-\alpha f/n} (e^{\alpha f} - 1) + (e^{-\alpha f/n} - 1) \cdot \Phi] \\
&= \sum_{i=1}^n p_i * e^{-\alpha f/n} \Phi_i (e^{\alpha f} - 1) + (e^{-\alpha f/n} - 1) \cdot p_i \cdot \Phi \\
&= \sum_{i=1}^n [p_i * e^{-\alpha f/n} (e^{\alpha f} - 1) + (e^{-\alpha f/n} - 1)] \Phi_i
\end{aligned} \tag{3.3}$$

Now,  $e^{(-\alpha f/n)} * (e^{\alpha f} - 1)$  can be approximated as follows:

$$\begin{aligned}
e^{(-\alpha f/n)} \cdot (e^{\alpha f} - 1) &\leq (1 - \alpha f/n + (\alpha f/n)^2) * (1 + \alpha f + (\alpha f)^2 - 1) \\
&\sim \alpha f + (\alpha f)^2 + O((\alpha f)^2/n) \\
e^{(-\alpha f/n)} \cdot (e^{\alpha f} - 1) &\lesssim (\alpha f + (\alpha f)^2)
\end{aligned}$$

Above, since,  $(\alpha f)^2/n$  is very small for large  $n$ , we have ignored the small terms. Similarly,  $(e^{-\alpha f/n} - 1) \lesssim -\alpha f/n$ . Hence, the expected change in  $\Phi$  can be given by:

$$\mathbb{E}[\Phi(t+1) - \Phi(t)|x(t)] \leq \sum_{i=1}^n [p_i * (\alpha f + (\alpha f)^2) - \alpha f/n] \Phi_i \tag{3.4}$$

Simplifying further and observing that  $\Phi_i$  decreases and  $p_i$  increases with increasing  $i$  from 1 to  $n$ , one gets the following Corollary.

**Corollary 3.2**  $\mathbb{E}[\Phi(t+1) - \Phi(t)|x(t)] \leq (\alpha f)^2 * \Phi/n$

**Proof** Since,  $p_i$  are increasing and  $\Phi_i$  are decreasing, the maximum value taken by RHS of equation (3.16) will be when  $p_i = 1/n$  for all  $i \in [1..n]$ . Simplifying, we get the result.

Similarly, the change in  $\Psi$  can be derived. For detailed proof refer to Appendix B.

**Lemma 3.3** *When an md-ball is thrown into an md-bin, the following inequality holds:*

$$\mathbb{E}[\Psi(t+1) - \Psi(t)|x(t)] \leq \sum_{i=1}^n [p_i * (-\alpha f + (\alpha f)^2) + \alpha f/n] \Psi_i \tag{3.5}$$

Further observing that  $p_i > 0$ , one gets the following Corollary.

**Corollary 3.4**  $\mathbb{E}[\Psi(t+1) - \Psi(t)|x(t)] \leq (\alpha f \cdot \Psi)/n$

In the next two lemmas, Lemma 3.5 and Lemma 3.6, we consider a reasonably balanced md-bins scenario. We show that for such cases, the expected potential decreases. Specifically, for  $s_{(n\gamma_2)} \leq 0$ , the expected value of  $\Phi$  decreases and for  $s_{(n\gamma_1)} \geq 0$ , the expected value of  $\Psi$  decreases.

**Lemma 3.5** *Let  $\Phi$  be defined as above. If  $s_{(n\gamma_2)}(t) < 0$  then,  $\mathbb{E}[\Phi(t+1)|x(t)] \leq (1 - \frac{\alpha f}{2n})\Phi$*

**Proof** From equation (3.16), we get,

$$\begin{aligned}
\mathbb{E}[\Phi(t+1) - \Phi(t)|x(t)] &\leq \sum_{i=1}^n (p_i * (\alpha f + (\alpha f)^2) - \alpha f/n) \cdot \Phi_i \\
&\leq \sum_{i < n\gamma_2} (p_i * (\alpha f + (\alpha f)^2)) \cdot \Phi_i - \frac{\alpha f \Phi}{n} + \sum_{i \geq n\gamma_2} p_i * (\alpha f + (\alpha f)^2) \cdot e^{-\alpha s_i}
\end{aligned} \tag{3.6}$$



Now, we need to upper bound the term  $\sum_{i < n\gamma_2} (p_i * (\alpha.f + (\alpha.f)^2) \cdot \frac{e^{\alpha.s_i}}{i})$ . Since  $p_i$  is non-decreasing and  $\Phi_i$  is non-increasing, the maximum value is achieved when  $e^{\alpha.s_i} (\sum_{i=1}^{n\gamma_2} 1/i) = \Phi$  for each  $i < n\gamma_2$ . Hence,  $e^{\alpha.s_i} = \frac{\Phi}{\ln(n\gamma_2)}$ . Hence, the maximum value is given as follows.

$$\begin{aligned} \sum_{i=1}^{n\gamma_2} p_i \Phi_i &\leq \frac{\Phi}{\ln(n\gamma_2)} * \sum_{i=1}^{n\gamma_2} [\frac{2i-1}{n^2} * \frac{1}{i}] \\ &\leq \frac{2\gamma_2 \Phi}{n \ln(n\gamma_2)} - \frac{\Phi}{n^2} \end{aligned} \quad (3.7)$$

Similarly, one can upper bound the term,  $\sum_{i \geq n\gamma_2} (p_i \frac{e^{-\alpha.s_i}}{i})$ . Since  $p_i$  is non-decreasing and  $\Phi_i$  is non-increasing, the maximum value is achieved when  $e^{-\alpha.s_i} (\sum_{i=(n\gamma_2)}^n 1/i) = \Phi_{(\geq n\gamma_2)}$  for each  $i \geq n\gamma_2$ . Hence,  $e^{-\alpha.s_i} = \frac{\Phi_{(\geq n\gamma_2)}}{\ln(1/\gamma_2)}$ . Hence, the required upper bound is given as follows.

$$\begin{aligned} \sum_{i=n\gamma_2}^n p_i \Phi_i &\leq \frac{\Phi}{\ln(1/\gamma_2)} * \sum_{i=n\gamma_2}^n [\frac{2i-1}{n^2} * \frac{1}{i}] \\ &\leq \frac{2(1-\gamma_2)\Phi}{n \ln(n)} \because \Phi_{(\geq n\gamma_2)} \leq \frac{\Phi}{\ln(n)} \end{aligned} \quad (3.8)$$

Thus, the expected change in  $\Phi$  can be computed, using equation (3.6) and the above bound, as follows:

$$\begin{aligned} \mathbb{E}[\Phi(t+1) - \Phi(t) | x(t)] &\leq (\alpha.f + (\alpha.f)^2) (\frac{2\gamma_2 \Phi}{n \ln(n\gamma_2)} - \frac{\Phi}{n^2}) - \alpha.f/n * \Phi + (\alpha.f + (\alpha.f)^2) * \frac{2(1-\gamma_2)\Phi}{n \ln(n)} \\ &\leq (\alpha.f)\Phi/n (\frac{2\gamma_2}{\ln(n\gamma_2)} - 1) + \frac{\epsilon^2 \gamma_2}{2n \ln(n\gamma_2)} + \alpha.f \frac{2(1-\gamma_2)\Phi}{n \ln(n)} \\ &\leq \frac{-\alpha.f\Phi}{2n} \end{aligned} \quad (3.9)$$

**Lemma 3.6** Let  $\Psi$  be defined as above. If  $s_{(n\gamma_1)}(t) \geq 0$  then,  $\mathbb{E}[\Psi(t+1) | x(t)] \leq (1 - \frac{\alpha.f}{8n})\Psi$

**Proof** The proof is similar to that of Lemma 3.5. See Appendix C for details of the proof.

Now, we consider the remaining cases and show that in case the load across the bins, at time  $t$ , is not reasonably balanced, then for  $s_{n\gamma_2} > 0$ , either  $\Psi$  dominates  $\Phi$  or  $\Gamma < c$ , where,  $c = \text{poly}(1/\epsilon)$ .

**Lemma 3.7** Let,  $s_{(n\gamma_2)} > 0$  and  $\mathbb{E}[\Delta\Phi | x(t)] \geq -\epsilon\Phi/4n$ . Then, either  $\Phi < \epsilon\gamma_1 * \Psi$ , or  $\Gamma < c$  for some  $c = \text{poly}(1/\epsilon)$ .

**Proof** From equation (3.16), we get:

$$\begin{aligned} \mathbb{E}[\Delta\Phi | x(t)] &\leq \sum_{i=1}^n (p_i * (\alpha.f + (\alpha.f)^2) - \alpha.f/n) \cdot \Phi_i \\ &\leq \sum_{i \leq n\gamma_3} (p_i * (\alpha.f + (\alpha.f)^2) - \alpha.f/n) \cdot \Phi_i + \sum_{i > n\gamma_3} (p_i * (\alpha.f + (\alpha.f)^2) - \alpha.f/n) * \Phi_i \\ &\leq [(\alpha.f + (\alpha.f)^2) * \sum_{i \leq (n\gamma_3)} \frac{2i-1}{n^2 i}] \cdot \frac{\Phi_{(\leq n\gamma_3)}}{\ln(n\gamma_3)} + \\ &\quad ((\sum_{i > (n\gamma_3)} \frac{2i-1}{n^2 i})) * (\alpha.f + (\alpha.f)^2) \frac{\Phi_{(> n\gamma_3)}}{\ln(1/\gamma_3)} - \frac{\alpha.f\Phi}{n} \\ &\leq \frac{\alpha.f\Phi_{\leq n\gamma_3}}{n} [\frac{2\gamma_3}{\ln(n\gamma_3)} - 1 - 1/n] + \frac{\alpha.f\Phi_{> n\gamma_3}}{n} [\frac{2\gamma_4}{\ln(1/\gamma_3)} - 1 - 1/n] \\ &\leq \frac{\alpha.f\Phi_{\leq n\gamma_3}}{n} [\frac{2\gamma_3 - \ln(n\gamma_3)}{\ln(n\gamma_3)}] + \frac{\alpha.f\Phi_{> n\gamma_3}}{n} [\frac{2\gamma_4 - \ln(1/\gamma_3)}{\ln(1/\gamma_3)}] \end{aligned} \quad (3.10)$$

Now, since  $\mathbb{E}[\Delta\Phi | x(t)] \geq -\alpha.f\Phi/2n$ , we get:  $\Phi \leq 4\Phi_{(> n\gamma_3)} [\frac{\gamma_4 \ln(n)}{\ln(n\gamma_3) \ln(1/\gamma_3)}]$ . Let,  $B = \sum_i \max(0, s_i)$ . Note,  $\sum_i s_i = 0$ , since for each dimension  $d$ , the update maintains that,  $\sum_{d=1}^D x_i^d(t) = 0$ . Further, because,  $s_{n\gamma_3} > 0$ ,  $\Phi_{(> n\gamma_3)} \leq \ln(1/\gamma_3) * e^{\frac{(\alpha.B)}{n\gamma_3}}$ . This implies that,  $\Phi \leq \frac{4\gamma_4 \ln(n) e^{\frac{(\alpha.B)}{n\gamma_3}}}{\ln(n\gamma_3)}$ .

Since,  $s_{(n\gamma_2)} > 0$ , so,  $\Psi \geq \ln(n\gamma_2) * e^{\frac{\alpha.B}{n\gamma_1}}$ . If  $\Phi < \epsilon\gamma_1 * \Psi$ , then we are done. Else,  $\Phi \geq \epsilon\gamma_1 * \Psi$ . This implies:

$$\frac{4\gamma_4 \ln(n) e^{\frac{\alpha.B}{n\gamma_3}}}{\ln(n\gamma_3)} \geq \Phi \geq \epsilon\gamma_1 * \Psi \geq \epsilon\gamma_1 \ln(n\gamma_2) * e^{\frac{\alpha.B}{n\gamma_1}}$$

Thus,  $e^{\alpha.B/n} \leq \left(\frac{4\gamma_4}{\epsilon\gamma_1}\right)^{\frac{\gamma_1\gamma_3}{(\gamma_3-\gamma_1)}}$ . So,  $\Gamma \leq \left(\frac{1+\theta}{\epsilon}\right) * \Phi \leq \frac{1+\theta}{\epsilon} * \frac{4\gamma_4 \ln(n)}{\ln(n\gamma_3)} * e^{\frac{\alpha.B}{n\gamma_3}} \leq \frac{1+\theta}{\epsilon} * \frac{4\gamma_4 \ln(n)}{\ln(n\gamma_3)} * \left(\frac{4\gamma_4}{\epsilon\gamma_1}\right)^{\frac{\gamma_1}{(\gamma_3-\gamma_1)}}$ . Hence,  $\Gamma < c$ , where,  $c = \text{poly}(1/\epsilon)$ .

In the Lemma below, we consider the case where the load across the bins at time,  $t$ , is not reasonably balanced, and  $s_{(n\gamma_1)} < 0$ . Here, we show that either  $\Phi$  dominates  $\Psi$  or the potential function is less than  $c$  for  $c = \text{poly}(1/\epsilon)$ .

**Lemma 3.8** *Let,  $s_{(n\gamma_1)} < 0$  and  $\mathbb{E}[\Delta\Psi|x(t)] \geq -\alpha f\Psi/8n$ . Then, either  $\Psi < \epsilon\gamma_1 * \Phi$ , or  $\Gamma < c \ln(n)$  for some  $c = \text{poly}(1/\epsilon)$ .*

**Proof** The proof is similar to that of Lemma 3.7. See Appendix D for details of the proof.

Now, we consider combinations of the cases considered so far and can show that the potential function,  $\Gamma$ , behaves as a super-martingale.

**Theorem 3.9** *For the potential function,  $\Gamma$ ,  $\mathbb{E}[\Gamma(t+1)|x(t)] \leq (1 - \frac{\epsilon}{24n(1+\epsilon\gamma_1)})\Gamma(t) + \frac{c \ln(n)}{n}$ , for constant  $c = \text{poly}(1/\epsilon)$ .*

**Proof** We consider the following cases on intervals of values for  $s_i$ .

- **Case 1:**  $s_{(n\gamma_1)} \geq 0$  and  $s_{(n\gamma_2)} \leq 0$ . Using, Lemma 3.5 and Lemma 3.6, we can immediately see that,  $\mathbb{E}[\Gamma(t+1)|x(t)] \leq (1 - \epsilon^2/16n)\Gamma(t)$  and hence, the result is also true.
- **Case 2:**  $s_{n\gamma_1} \geq s_{n\gamma_2} > 0$ . This represents a high load imbalance across the bins. In some cases,  $\Phi$  may grow but the asymmetry in the load implies that  $\Gamma$  is dominated by  $\Psi$ . Thus, the decrease in  $\Psi$  offsets the increase in  $\Phi$  and hence the expected change in  $\Gamma$  is negative.  
Specifically, if  $\mathbb{E}[\Delta\Phi|x] \leq \frac{-\alpha f\Phi}{2n}$ , then using Lemma 3.6 we get that  $\mathbb{E}[\Gamma(t+1)|x(t)] \leq (1 - \alpha f/8n)\Gamma(t)$ ; else we consider the following two cases:

- **Case 2a:**  $\Phi < \epsilon\gamma_1 * \Psi$ . Here, using Lemma 3.6 and Corollary 3.2, we get:

$$\begin{aligned} \mathbb{E}[\Delta\Gamma|x] &= \mathbb{E}[\Delta\Phi|x] + \mathbb{E}[\Delta\Psi|x] \\ &\leq \frac{(\alpha.f)^2}{n}.\Phi - \frac{\alpha f}{8n} * \Psi \\ &\leq -\frac{\epsilon}{24n}.\Psi \\ &\leq -\frac{\epsilon}{24n(1+\epsilon\gamma_1)}\Gamma \end{aligned}$$

- **Case 2b:**  $\Gamma < c \ln(n)$ . Here, using Corollary 3.4 and Corollary 3.2, we get:

$$\mathbb{E}[\Delta\Gamma|x] \leq \alpha.f/n * \Gamma \leq \frac{c\alpha.f \ln(n)}{n}$$

But,  $c \ln(n)/n - ((\epsilon/4n) * \Gamma) \geq c \ln(n)/n(1 - \epsilon/4) \geq c \ln(n)/n(1 - \epsilon/2) \geq \frac{c\alpha.f \ln(n)}{n}$ . Hence,  $\mathbb{E}[\Delta\Gamma|x] \leq -\frac{\epsilon\Gamma}{4n} + \frac{c \ln(n)}{n}$ .

- **Case 3:**  $s_{n\gamma_2} \leq s_{n\gamma_1} < 0$ . Here, if  $\mathbb{E}[\Delta\Psi|x] \leq \frac{-\epsilon}{16n}\Psi$ , then using Lemma 3.5, we get that  $\mathbb{E}[\Gamma(t+1)|x(t)] \leq (1 - \epsilon/16n)\Gamma(t)$ ; else we consider the following two cases:

- **Case 3a:**  $\Psi < \epsilon\gamma_1 * \Phi$ . Here, using Lemma 3.5 and Corollary 3.4, we get:

$$\begin{aligned} \mathbb{E}[\Delta\Gamma|x] &= \mathbb{E}[\Delta\Phi|x] + \mathbb{E}[\Delta\Psi|x] \\ &\leq -(\epsilon/4n)\Phi + \alpha.f/n * \Psi \\ &\leq -(\epsilon/4n).\Phi + (\gamma_1\epsilon^2/2n)\Phi \\ &\leq \frac{-\epsilon}{8n}\Phi \\ &\leq \frac{-\epsilon}{(8n(1+\epsilon\gamma_1))} * \Gamma \end{aligned}$$

– **Case 3b:**  $\Gamma < c \ln(n)$ . Here, using Corollary 3.2 and Corollary 3.4, we get:

$$\mathbb{E}[\Delta\Gamma|x] = \alpha \cdot f/n * \Gamma \leq c\alpha \cdot f \ln(n)/n$$

Hence, this case follows similarly as *Case 2b* above.

Now, we can prove using induction that the expected value of  $\Gamma$  remains bounded.

**Theorem 3.10** For any time  $t \geq 0$ ,  $\mathbb{E}[\Gamma(t)] \leq \frac{24c(1+\epsilon\gamma_1)}{\epsilon} \ln(n)$

**Proof** Using induction we can prove this claim. For  $t = 0$ , it is trivially true since  $\Gamma(0) = 2 \ln(n)$ . Using Theorem 3.9, we get:

$$\begin{aligned} \mathbb{E}[\Gamma(t+1)] &= E[E[\Gamma(t+1)|\Gamma(t)]] \\ &\leq \mathbb{E}[(1 - \frac{\epsilon}{24n(1+\epsilon\gamma_1)})\Gamma(t) + \frac{c \ln(n)}{n}] \\ &\leq \frac{24c(1+\epsilon\gamma_1)}{\epsilon} \ln(n) - c \frac{\ln(n)}{n} + \frac{c \ln(n)}{n} \\ &\leq \frac{24c(1+\epsilon\gamma_1)}{\epsilon} \ln(n) \end{aligned}$$

Now, we can upper bound the gap across all the  $D$  dimensions across all the  $n$  md-bins. This gap is defined as follows:

$$Gap(t) = \max_{d=1}^D [\max_{i=1}^n x_i^d] \quad (3.11)$$

**Theorem 3.11 Fixed  $f$  Case:** Using the bias in the probability distribution in favor of lightly loaded md-bins as given by the  $d$ -choice algorithm, and assuming that  $f$  dimensions are exactly populated in each md-ball with uniform distribution of  $f$  dimensions over  $D$ , the expected and probabilistic upper bound on the gap (maximum dimensional gap) across the multidimensional bins is given as follows. Let,  $\delta = \frac{24c(1+\epsilon\gamma_1)}{\epsilon}$ , then:

$$\begin{aligned} E[Gap(t)] &\leq 2 \log \log(n)/\epsilon + 2 \log \log(\delta)/\epsilon \\ Pr[Gap(t) > (\frac{mf}{nD})^{1/2+\zeta} * (4 \log \log(n)/\epsilon + 4 \log(\log \delta)/\epsilon)] &\leq D/mf \end{aligned}$$

**Proof** Let,  $a$  be the winning md-bin and  $m$  be the winning dimension that represents  $Gap(t)$ . Now, from Theorem 3.10, we get,  $\mathbb{E}[e^{\alpha \cdot s_a}] \leq \delta \log(n)$ . So,  $\mathbb{E}[e^{\alpha \cdot (x_a^m + \sum_{d \neq m} x_a^d)}] \leq \delta \log(n)$ . Let,  $y_a$  denote the gap as measured by the number of md-balls in bin  $a$  minus the average number of balls across the bins. Then,

$$\begin{aligned} E[s_a] &\leq 1/\alpha * \log \log(n) + 1/\alpha * \log \log(\delta) \\ \Rightarrow E[s_a] &\leq 2f \log \log(n)/\epsilon + O(2f \log \log(\delta)/\epsilon) \\ \Rightarrow f \cdot E[y_a] &\leq 2f \log \log(n)/\epsilon + O(2f \log \log(\delta)/\epsilon) \\ \Rightarrow E[y_a] &\leq 2 \log \log(n)/\epsilon + O(2 \log \log(\delta)/\epsilon) \end{aligned} \quad (3.12)$$

The third inequality uses the fact that each ball has exactly  $f$  populated dimensions. Since the  $f$  dimensions are chosen uniformly and randomly from  $D$  dimensions, the expected gap in any dimension (and hence the winning dimension with the maximum gap) is bounded by  $O(\log \log(n))$ . Now, consider the case of a non-uniform distribution, where we assume that each dimension is chosen with probability at most  $\kappa_2$  in each md-ball and each md-ball still has fixed  $f$  populated dimensions. Here, one can see that the expected gap can be bounded by  $O(\kappa_2 \log \log(n))$ .

Now, the  $Pr[s_a > 4f \log \log(n)/\epsilon + 4f \log \log(\delta)/\epsilon] \leq Pr[\Gamma(t) \geq nE[\Gamma(t)]] \leq 1/n$  (using Markov's Inequality); where  $s_a = \sum_{d=1}^D x_a^d$ . Further, the probability that within a single md-bin, a particular dimension has more than the expected number of 1s, can be given by the Chernoff Bound as follows. Let  $m/n$  balls be thrown into an md-bin. The number of ones in any dimension follows a Binomial distribution,  $B(m/n, f/D)$ . Using Chernoff Bound, and assuming  $t = (\frac{mf}{nD})^{1/2+\zeta}$ , we have:

$$\begin{aligned} Pr[B(m/n, f/D) > (mf/nD + t)] &\leq (\frac{mf/nD}{mf/nD + t})^{mf/nD + t} * e^t \\ \Rightarrow Pr[B(m/n, f/D) > (mf/nD + t)] &\leq nD/mf \end{aligned} \quad (3.13)$$

Hence,  $Pr[y_a > (\frac{mf}{nD})^{1/2+\zeta} * (4 \log \log(n)/\epsilon + 4 \log \log(\delta)/\epsilon)] \leq 1/n * nD/mf = D/mf$ .

### 3.3 Lower Bound for Unweighted Case

We can show that the expected upper bound, for fixed  $f$  case with uniform distribution, proved in section 3.2 is tight to within  $f/D$  factor. Consider the case when,  $m$  balls are thrown into  $n$  bins, using the  $d$  choice process. The expected dimensional sum load per bin is  $fm/n$ . Berenbrink et.al. [4] show that when  $m \gg n$  balls are thrown using the  $d$  choice process into  $n$  bins, then the load of the most loaded bin is at least  $O(\ln \ln(n))$  balls more than the average  $m/n$ . Thus, for  $md$ -balls the sum load of the most loaded  $md$ -bin is at least  $\Omega(f \ln \ln(n) + fm/n)$ . Since, each ball has  $f$  populated dimensions, hence, there are at least  $\Omega(\ln \ln(n) + m/n)$  balls in this max sum load bin. Since, in each ball  $f$  dimensions are uniformly distributed over  $D$  dimensions, there exists a dimension whose load is at least  $\Omega(f \ln \ln(n)/D)$  more than the average  $mf/nD$ . Hence, the lower bound is  $O(f \ln \ln(n)/D)$ .

### 3.4 Parallel Multidimensional Balls & Bins: Unweighted Case

Consider the following parallel  $d$ -choice process. Let  $m$  balls be thrown in parallel using  $d$ -choice process into  $n$  bins. In each round, a bin sends the (ball's) rank to the ball with the lowest ID. The ball chooses the bin (out of  $d$  bins it selected) that gives the lowest rank. It can be shown that this parallel process produces exactly the same distribution of balls in the bins as a sequential *Greedy with Ties* process [1]. In the sequential *Greedy with Ties* process, when there are multiple bins with same lowest load, all of these bins get the ball. Using the potential function analysis as above, we can show that the gap in this case, can also be bounded by  $O(\log \log(n))$ . We provide an overview of the proof below. Consider the change in  $\Phi(t)$  (also refers to  $\Phi$  by default) and  $\Psi(t)$  (also refers to  $\Psi$  by default) separately when a ball is thrown with the given probability distribution.

**Lemma 3.12** *When an  $md$ -ball is thrown into an  $md$ -bin, the following inequality holds:*

$$\mathbb{E}[\Phi(t+1) - \Phi(t)|x(t)] \leq \sum_{i=1}^n [d * p_i * (\alpha.f + (\alpha.f)^2) - d\alpha.f/n].e^{\alpha.s_i} \quad (3.14)$$

**Proof** In the Greedy with Ties process, some number (less than  $d$ ) of bins each with the same load (and hence belonging to the same equi-load group) can get the (*replicated*) ball. In the worst case all the  $d$  randomly selected bins, chosen by the ball, have the same load and hence get the  $md$ -ball. All of these  $md$ -bins, then move to the previous equi-load group or a new equi-load group is created. Let  $\Delta$  be the expected change in  $\Phi$  when the ball is put in a certain number (less than  $d$ ) of bins. If one of these bins is  $i$ , then,  $r_i(t+1) = s_i + f(1 - d/n)$ . For bins  $j \neq i$ , that do not get the  $md$ -ball,  $r_j(t+1) = s_j(t) - df/n$ . The new values i.e.  $s(t+1)$  are obtained by sorting  $r(t+1)$  and  $\Phi(s) = \Phi(r)$ . Thus, the expected contribution of bin,  $i$ , to  $\Delta$  is given as follows:

$$\mathbb{E}\left[\frac{e^{\alpha.(s_i + f(1-d/n))}}{i}\right] - \frac{e^{\alpha.s_i}}{i} = \frac{e^{\alpha.s_i}}{i} [e^{\alpha.f(1-d/n)} - 1]$$

Similarly, the expected contribution of bin (that does not get the ball),  $j$  ( $j \neq i$ ) to  $\Delta$  is given as:

$$\mathbb{E}\left[\frac{e^{\alpha.(s_j - df/n)}}{j}\right] - \frac{e^{\alpha.s_j}}{j} = \frac{e^{\alpha.s_j}}{j} [e^{-\alpha.df/n} - 1]$$

Assuming that the bins that get the replicated ball are  $i_1, i_2, \dots, i_d$ ,  $\Delta$  is given as follows:

$$\begin{aligned} \Delta &= d\Phi_i [e^{\alpha.f(1-d/n)} - 1] + \sum_{j \neq (i_1, i_2, \dots, i_d)} \Phi_j (e^{-\alpha.df/n} - 1) \\ &= d\Phi_i e^{-\alpha.df/n} (e^{\alpha.f} - 1) + (e^{-\alpha.df/n} - 1) \cdot \Phi \end{aligned}$$

Thus, we get the overall expected change in  $\Phi$  as follows:

$$\begin{aligned} \mathbb{E}[\Phi(t+1) - \Phi(t)|x(t)] &= \sum_{i=1}^n p_i * \Delta \\ &= \sum_{i=1}^n p_i * [d\Phi_i e^{-\alpha.df/n} (e^{\alpha.f} - 1) + (e^{-\alpha.df/n} - 1) \cdot \Phi] \\ &= \sum_{i=1}^n p_i * de^{-\alpha.df/n} \Phi_i (e^{\alpha.f} - 1) + (e^{-\alpha.df/n} - 1) \cdot p_i \cdot \Phi \\ &= \sum_{i=1}^n [p_i * de^{-\alpha.df/n} (e^{\alpha.f} - 1) + (e^{-\alpha.df/n} - 1)] \Phi_i \end{aligned} \quad (3.15)$$

Now,  $e^{(-\alpha \cdot f d/n)} * (e^{\alpha \cdot f} - 1)$  can be approximated as follows:

$$\begin{aligned} e^{(-\alpha \cdot f d/n)} \cdot (e^{\alpha \cdot f} - 1) &\leq (1 - \alpha \cdot f d/n + (\alpha \cdot f d/n)^2) * (1 + \alpha \cdot f + (\alpha \cdot f)^2 - 1) \\ &\sim \alpha \cdot f + (\alpha \cdot f)^2 + O((\alpha \cdot f d/n)^2/n) \\ \Rightarrow e^{(-\alpha \cdot f d/n)} \cdot (e^{\alpha \cdot f} - 1) &\lesssim (\alpha \cdot f + (\alpha \cdot f)^2) \end{aligned}$$

Above, since,  $(\alpha \cdot f d)^2/n$  is very small for large  $n$ , we have ignored the small terms. Similarly,  $(e^{-\alpha \cdot f d/n} - 1) \lesssim -\alpha \cdot f d/n$ . Hence, the expected change in  $\Phi$  can be given by:

$$\mathbb{E}[\Phi(t+1) - \Phi(t)|x(t)] \leq \sum_{i=1}^n [p_i * d(\alpha \cdot f + (\alpha \cdot f)^2) - \alpha \cdot f d/n] \Phi_i \quad (3.16)$$

Similarly, one can show the following.

**Lemma 3.13** *When an md-ball is thrown into an md-bin, the following inequality holds:*

$$\mathbb{E}[\Psi(t+1) - \Psi(t)|x(t)] \leq \sum_{i=1}^n [d * p_i * (-\alpha \cdot f + (\alpha \cdot f)^2) + d\alpha \cdot f/n] \cdot e^{-\alpha \cdot s_i} \quad (3.17)$$

Following similar lines of proof as for the sequential multidimensional case, one can hence show that:

**Theorem 3.14** *For any time  $t \geq 0$ ,  $\mathbb{E}[\Gamma(t)] \leq \frac{24cd(1+\epsilon\gamma_1)}{\epsilon} \log(n)$*

Thus, this parallel balls and bins process with  $m \gg n$  balls and  $n$  bins, takes  $O(\frac{m}{n} + \log \log(n))$  rounds and results in maximum bin load  $O(\frac{m}{n} + \log \log(n))$  resulting in upper bound on the gap of  $O(\log \log(n))$ . Hence, one can derive that the gap (using Theorem 3.11) for the multidimensional parallel scenario is also bounded by  $O((mf/D)^{(1/2+\zeta)} \log \log(n))$  with high probability.

### 3.5 Upper Bound On Gap: Weighted Case

Here, we consider the case when the multidimensional balls have variable number of populated dimensions,  $f$ . The sum of dimensional load in an md-ball,  $f$ , is thus a random variable. We assume that the distribution for  $f$  has a finite second moment and average value,  $f^*$ . For this distribution, we assume that there is a  $\lambda > 0$  such that the moment generating function  $M[\lambda] = E[e^{\lambda \cdot f}] < \infty$ . Note that  $M''(z) = E[f^2 e^{zf}] \leq \sqrt{E[f^4]E[e^{2zf}]}$ . The above assumption implies that there is a  $S \geq 1$ , such that for every  $|z| < \lambda/2$  it holds that  $M''(z) < 2S$ . Our analysis below is primarily for integer valued  $f$  and for the multidimensional case. However, it can be easily seen that similar analysis holds for scalar balls and bins with real valued weight per ball  $W$  and  $E[W] = 1$  (still assuming that the distribution of  $W$  has finite second moment).

The weighted case is more challenging than the unweighted case, since we have to carefully consider the change in the rank of a bin when an md-ball of total dimensional load (weight)  $f$  falls in it, as the change in rank could increase the potential by a large amount. Thus, the potential function used in section 3.2 might not work in this case and we need to devise a new one. Assume that  $\epsilon \leq 1/4$ . Further, let  $\alpha = \min(\frac{\epsilon}{6S}, \frac{2}{\lambda}, \frac{\epsilon}{2f^*})$ . Define the following potential functions over the bins:

$$\begin{aligned} \Phi(t) &= \Phi(s(t)) = \sum_{i=1}^n \frac{e^{\alpha \cdot s_i}}{n^2 + i} \\ \Psi(t) &= \Psi(s(t)) = \sum_{i=1}^n \frac{e^{-\alpha \cdot s_i}}{n^2 + n - i + 1} \\ \Gamma(t) &= \Gamma(s(t)) = \Phi(t) + \Psi(t) \\ &= \sum_{i=1}^n \left[ \frac{e^{\alpha \cdot s_i}}{n^2 + i} + \frac{e^{-\alpha \cdot s_i}}{n^2 + n - i + 1} \right] \end{aligned} \quad (3.18)$$

where,  $s_i(t) = \sum_{d=1}^D x_i^d(t)$

In the beginning, each dimension for each bin has 0 weight, thus  $s_i = 0, \forall i$  and hence,  $\Gamma(0) \leq 2(n/(n^2 + 1)) \leq 2/n$ . We show that if  $\Gamma(x(t)) \geq a/n$  for some  $a > 0$ , then  $\mathbb{E}[\Gamma(t+1)|x(t)] \leq (1 - \frac{\alpha \cdot f^*}{16n(1+\epsilon\gamma_1)}) * \Gamma(t)$ . This helps in demonstrating that for every given  $t$ ,  $\mathbb{E}[\Gamma(t)] \in O(1/n)$ . This implies that the maximum gap is  $O(\log(n))$  w.h.p.

First, consider the change in  $\Phi(t)$  (also refers to  $\Phi$  by default) and  $\Psi(t)$  (also refers to  $\Psi$  by default) separately when a ball is thrown with the given probability distribution. Let there be constants,  $0 < \gamma_1 < \gamma_2 < 1/2 < \gamma_4 < \gamma_3$ , such that  $\gamma_2 + \gamma_3 > 1$  and  $\gamma_1 + \gamma_4 < 1$  and  $\gamma_2 < 7/16$

**Lemma 3.15** *When an md-ball is thrown into an md-bin, the following inequality holds:*

$$\mathbb{E}[\Phi(t+1) - \Phi(t)|x(t)] \leq \sum_{i=1}^n [p_i * (\alpha.f^* + 1/n + S\alpha^2) - \alpha.f^*/n].e^{\alpha.s_i} \quad (3.19)$$

**Proof** Let  $\Delta_i$  be the expected change in  $\Phi$  if the ball is put in bin,  $i$ . So,  $r_i(t+1) = s_i + f(1 - 1/n)$ ; and for  $j \neq i$ ,  $r_j(t+1) = s_j(t) - f/n$ . The new values i.e.  $s(t+1)$  are obtained by sorting  $r(t+1)$  and  $\Phi(s) = \Phi(r)$ . When, an md-ball is committed to bin  $i$ , then it jumps to an index  $i_{new}$  which is less than or equal to  $i$  in the new bin order. Thus, the expected contribution of bin,  $i$ , to  $\Delta_i$  is given as follows:

$$\begin{aligned} \mathbb{E}\left[\frac{e^{\alpha.(s_i + f(1-1/n))}}{n^2 + i_{new}}\right] - \frac{e^{\alpha.s_i}}{n^2 + i} &\leq \frac{e^{\alpha.s_i}}{n^2 + i} \left[\frac{M(\alpha(1 - 1/n))(n^2 + n)}{n^2 + 1} - 1\right] \\ &\leq \Phi_i[(M(0) + M'(0).\alpha(1 - 1/n) + M''(0)(\alpha(1 - 1/n))^2)(1 + 1/n) - 1] \\ &\leq \Phi_i[(1 + f^*\alpha(1 - 1/n) + S\alpha^2)(1 + 1/n) - 1] \\ &\because M(0) = 1, M'(0) = E(f) = f^*, M''(0) \leq 2S \\ &\leq \Phi_i[f^*\alpha(1 - 1/n) + S\alpha^2 + (1 + f^*\alpha(1 - 1/n) + S\alpha^2)/n] \\ &\leq \Phi_i[f^*\alpha + 1/n + S\alpha^2] \end{aligned} \quad (3.20)$$

The bins that were at index  $j \in [i_{new}..(i-1)]$ , shift right by one position and hence the expected contribution of such a bin,  $j$  to  $\Delta_i$  is given as:

$$\begin{aligned} \mathbb{E}\left[\frac{e^{\alpha.(s_j - f/n)}}{n^2 + j + 1}\right] - \frac{e^{\alpha.s_j}}{n^2 + j} &\leq \frac{e^{\alpha.s_j}}{n^2 + j} [M(-\alpha/n) - 1] \\ &\leq \Phi_j[M(0) + M'(0)(-\alpha/n) + M''(0)\frac{\alpha^2}{2n^2}] \\ &\leq \Phi_j\left[\frac{-f^*\alpha}{n} + \frac{S\alpha^2}{n^2}\right] \end{aligned} \quad (3.21)$$

For all other bins, their rank does not change in the new bin order, hence, their expected contribution to  $\Delta_i$  is given as:

$$\begin{aligned} \mathbb{E}\left[\frac{e^{\alpha.(s_j - f/n)}}{n^2 + j}\right] - \frac{e^{\alpha.s_j}}{n^2 + j} &= \Phi_j\left[\frac{-f^*\alpha}{n} + \frac{S\alpha^2}{n^2}\right] \end{aligned} \quad (3.22)$$

Using equations (3.20), (3.21) and (3.22),  $\Delta_i$  is given as follows:

$$\Delta_i = (-\alpha.f^*/n + 1/n + S\alpha^2)\Phi_i + \frac{\alpha.f^*\Phi}{n}$$

Hence, the expected change in  $\Phi$  can be given by:

$$\mathbb{E}[\Phi(t+1) - \Phi(t)|x(t)] \leq \sum_{i=1}^n [p_i * (\alpha.f^* + 1/n + S\alpha^2) - \alpha.f^*/n]\Phi_i \quad (3.23)$$

Simplifying further and observing that  $\Phi_i$  decreases and  $p_i$  increases with increasing  $i$  from 1 to  $n$ , one gets the following Corollary.

**Corollary 3.16**  $\mathbb{E}[\Phi(t+1) - \Phi(t)|x(t)] \leq (\alpha.f^* + 2S\alpha^2)\frac{\Phi}{n}$

**Proof** Since,  $p_i$  are increasing and  $\Phi_i$  are decreasing, the maximum value taken by RHS of equation (3.23) will be when  $e^{\alpha \cdot s_i} * \sum_{i=1}^n \frac{1}{(n^2+i)} = \Phi$ . Thus,  $e^{\alpha \cdot s_i} = n\Phi$ . Hence,

$$\begin{aligned} \sum_{i=1}^n p_i \Phi_i &\leq n\Phi * \sum_{i=1}^n \frac{2i-1}{n^2(n^2+i)} \\ &\leq \frac{\Phi}{n} * \frac{2n-1}{n+1} \end{aligned} \quad (3.24)$$

Using, equation (3.23), we get:

$$\begin{aligned} \mathbb{E}[\Phi(t+1) - \Phi(t)|x(t)] &\leq (\alpha \cdot f^*(1 + 1/n) + 1/n + S\alpha^2) * \frac{(2n-1)\Phi}{n(n+1)} - \frac{\alpha \cdot f^* \Phi}{n} \\ &\leq (\alpha \cdot f^* + 2S\alpha^2) \frac{\Phi}{n} \end{aligned} \quad (3.25)$$

Similarly, the change in  $\Psi$  can be derived. For detailed proof refer to Appendix E.

**Lemma 3.17** When an md-ball is thrown into an md-bin, the following inequality holds:

$$\mathbb{E}[\Psi(t+1) - \Psi(t)|x(t)] \leq \sum_{i=1}^n [p_i * (-\alpha \cdot f^* + \frac{S\alpha^2}{n^2} + \alpha \cdot f^*/n) \Psi_i] \quad (3.26)$$

Further observing that  $p_i > 0$ , one gets the following Corollary.

**Corollary 3.18**  $\mathbb{E}[\Psi(t+1) - \Psi(t)|x(t)] \leq (\alpha \cdot f^* \Psi)/n$

In the next two lemmas, Lemma 3.19 and Lemma 3.20, we consider a reasonably balanced md-bins scenario. We show that for such cases, the expected potential decreases. Specifically, for  $s_{(n\gamma_2)} \leq 0$ , the expected value of  $\Phi$  decreases and for  $s_{(n\gamma_1)} \geq 0$ , the expected value of  $\Psi$  decreases.

**Lemma 3.19** Let  $\Phi$  be defined as above. If  $s_{(n\gamma_2)}(t) < 0$  then,  $\mathbb{E}[\Phi(t+1)|x(t)] \leq (1 - \frac{\alpha f^*}{8n})\Phi$

**Proof** From equation (3.23), we get,

$$\begin{aligned} \mathbb{E}[\Phi(t+1) - \Phi(t)|x(t)] &\leq \sum_{i=1}^n (p_i * (\alpha f^* + S(\alpha f)^2) - \alpha f^*/n) \cdot \Phi_i \\ &\leq \sum_{i < n\gamma_2} (p_i * (\alpha f^* + S(\alpha)^2)) \cdot \Phi_i - \frac{\alpha f^* \Phi}{n} + \sum_{i \geq n\gamma_2} p_i * (\alpha f^* + S(\alpha)^2) \cdot e^{-\alpha s_i} \end{aligned} \quad (3.27)$$

Now, we need to upper bound the term  $\sum_{i < n\gamma_2} (p_i * (\alpha \cdot f^* + (\alpha)^2) \cdot \frac{e^{\alpha \cdot s_i}}{n^2+i})$ . Since  $p_i$  is non-decreasing and  $\Phi_i$  is non-increasing, the maximum value is achieved when  $e^{\alpha s_i} \sum_{i=1}^{n\gamma_2} \frac{1}{(n^2+i)} = \Phi$  for each  $i < n\gamma_2$ . Hence,  $e^{\alpha s_i} = \frac{\Phi(n+\gamma_2)}{\gamma_2}$ . Hence, the maximum value is given as follows.

$$\begin{aligned} \sum_{i=1}^{n\gamma_2} p_i \Phi_i &\leq \frac{\Phi(n+\gamma_2)}{\gamma_2} * \sum_{i=1}^{n\gamma_2} [\frac{2i-1}{n^2} * \frac{1}{n^2+i}] \\ &\leq \frac{(n+\gamma_2)\Phi}{n^2\gamma_2} * \frac{\gamma_2(2n\gamma_2-1)}{(n+\gamma_2)} \\ &\leq \frac{(2n\gamma_2-1)\Phi}{n^2} \end{aligned} \quad (3.28)$$

Similarly, one can upper bound the term,  $\sum_{i \geq n\gamma_2} (p_i \frac{e^{\alpha \cdot s_i}}{n^2+i})$ . Since  $p_i$  is non-decreasing and  $\Phi_i$  is non-increasing, the maximum value is achieved when  $e^{\alpha s_i} (\sum_{i=(n\gamma_2)}^n \frac{1}{n^2+i}) = \Phi_{(\geq n\gamma_2)}$  for each  $i \geq n\gamma_2$ . Hence,  $e^{\alpha s_i} = \frac{\Phi_{(\geq n\gamma_2)}(n+\gamma_2)}{(1-\gamma_2)}$ . Thus, the expected change in  $\Phi$  can be computed, using equation (3.27) and the above bound, as follows:

$$\begin{aligned} \mathbb{E}[\Delta\Phi|x(t)] &\leq (\alpha \cdot f^* + S\alpha^2) \frac{(2n\gamma_2-1)\Phi}{n^2} - \frac{\alpha \cdot f^*}{n} * \Phi + (\alpha \cdot f^* + S\alpha^2) * \frac{(2n-1)(n+\gamma_2)\Phi_{(\geq n\gamma_2)}}{n^2(n+1)} \\ &\leq \frac{2\alpha \cdot f^* \gamma_2 \Phi}{n} - \frac{\alpha \cdot f^* \Phi}{n} \\ &\leq \frac{(2\gamma_2-1)\alpha \cdot f^* \Phi}{n} \\ &\leq \frac{-\alpha \cdot f^* \Phi}{8n} \because \gamma_2 < (1/2 - 1/16) \end{aligned} \quad (3.29)$$

**Lemma 3.20** Let  $\Psi$  be defined as above. If  $s_{(n\gamma_1)}(t) \geq 0$  then,  $\mathbb{E}[\Psi(t+1)|x(t)] \leq (1 - \frac{\alpha.f^*}{2n})\Psi$

**Proof** The proof is similar to that of Lemma 3.19. See Appendix F for details of the proof.

Now, we consider the remaining cases and show that in case the load across the bins, at time  $t$ , is not reasonably balanced, then for  $s_{n\gamma_2} > 0$ , either  $\Psi$  dominates  $\Phi$  or  $\Gamma < c/n$ , where,  $c = \text{poly}(1/\epsilon)$ .

**Lemma 3.21** Let,  $s_{(n\gamma_2)} \geq 0$  and  $\mathbb{E}[\Delta\Phi|x(t)] \geq -\alpha.f^*\Phi/8n$ . Then, either  $\Phi < \epsilon\gamma_1 * \Psi$ , or  $\Gamma < \frac{c}{n}$  for some  $c = \text{poly}(1/\epsilon)$ .

**Proof** From equation (3.16), we get:

$$\begin{aligned}
\mathbb{E}[\Delta\Phi|x(t)] &\leq \sum_{i=1}^n (p_i * (\alpha.f^* + S\alpha^2) - \alpha.f^*/n) \cdot \Phi_i \\
&\leq \sum_{i \leq n\gamma_3} (p_i * (\alpha.f^* + S\alpha^2) - \alpha.f^*/n) \cdot \Phi_i + \sum_{i > n\gamma_3} (p_i * (\alpha.f^* + S\alpha^2) - \alpha.f^*/n) * \Phi_i \\
&\leq \frac{n\Phi_{(\leq n\gamma_3)}}{\gamma_3} (\alpha.f^* + S\alpha^2) * \sum_{i \leq (n\gamma_3)} \frac{2i-1}{n^2(n^2+i)} + \\
&\quad (\alpha.f^* + S\alpha^2) \frac{n\Phi_{(>n\gamma_3)}}{1-\gamma_3} * \sum_{i > (n\gamma_3)} \frac{2i-1}{n^2(n^2+i)} - \frac{\alpha.f^*\Phi}{n} \\
&\leq \frac{\alpha.f^*\Phi_{\leq n\gamma_3}}{n} \frac{(2n\gamma_3-1)}{n+\gamma_3} + \frac{\alpha.f^*\Phi_{>n\gamma_3}}{1-\gamma_3} \frac{\gamma_3}{n+1} - \frac{\alpha.f^*\Phi}{n} \\
&\leq \frac{\alpha.f^*\Phi}{n} \left( \frac{2\gamma_3}{n+\gamma_3} - 1 \right) + \alpha.f^*\Phi_{>n\gamma_3} \left[ -\frac{2n\gamma_3-1}{n(n+\gamma_3)} + \frac{\gamma_3}{(1-\gamma_3)(n+1)} \right]
\end{aligned} \tag{3.30}$$

Now, since  $\mathbb{E}[\Delta\Phi|x(t)] \geq -\alpha.f\Phi/8n$ , we get:  $\Phi \leq 4n\Phi_{(>n\gamma_3)} \left[ \frac{(n-2)\gamma_3+1}{(n+1)(1-\gamma_3)(4n+3\gamma_3)} \right]$ . Let,  $B = \sum_i \max(0, s_i)$ . Note,  $\sum_i s_i = 0$ , since for each dimension  $d$ , the update maintains that,  $\sum_{d=1}^D x_i^d(t) = 0$ . Further, because,  $s_{n\gamma_3} > 0$ ,  $\Phi_{(>n\gamma_3)} \leq \frac{1-\gamma_3}{n+\gamma_3} * e^{\frac{\alpha.B}{n\gamma_3}}$ . This implies that,  $\Phi \leq \frac{4(n-2)\gamma_3 e^{\frac{\alpha.B}{n\gamma_3}}}{(4n+3\gamma_3)(n+\gamma_3)}$ .

Since,  $s_{(n\gamma_2)} > 0$ , so,  $\Psi \geq \frac{n}{\gamma_2} * e^{\frac{\alpha.B}{(n-n\gamma_2)}}$ . If  $\Phi < \epsilon\gamma_1 * \Psi$ , then we are done. Else,  $\Phi \geq \epsilon\gamma_1 * \Psi$ . This implies:

$$e^{\frac{\alpha.B}{n\gamma_3}} \frac{4(n-2)\gamma_3}{(4n+3\gamma_3)(n+\gamma_3)} \geq \Phi \geq \epsilon\gamma_1 * \Psi \geq \frac{n\epsilon\gamma_1}{\gamma_2} * e^{\frac{\alpha.B}{(n-n\gamma_2)}}$$

Thus,  $e^{\alpha.B/n} \leq \left( \frac{4(n-2)\gamma_3\gamma_2}{n\epsilon\gamma_1(4n+3\gamma_3)(n+\gamma_3)} \right)^{\frac{(1-\gamma_2)\gamma_3}{(\gamma_3+\gamma_2-1)}}$ . So,  $\Gamma \leq \frac{1+\theta}{\epsilon} * \Phi \leq \frac{1+\theta}{\epsilon} * \frac{4(n-2)\gamma_3}{(4n+3\gamma_3)(n+\gamma_3)} * e^{\frac{\alpha.B}{n\gamma_3}}$

Hence,  $\Gamma < c/n$ , where,  $c = \text{poly}(1/\epsilon)$ .

In the Lemma below, we consider the case where the load across the bins at time,  $t$ , is not reasonably balanced, and  $s_{(n\gamma_1)} < 0$ . Here, we show that either  $\Phi$  dominates  $\Psi$  or the potential function is less than  $c/n$  for  $c = \text{poly}(1/\epsilon)$ .

**Lemma 3.22** Let,  $s_{(n\gamma_1)} < 0$  and  $\mathbb{E}[\Delta\Psi|x(t)] \geq -\alpha.f^*\Psi/2n$ . Then, either  $\Psi < \epsilon\gamma_1 * \Phi$ , or  $\Gamma < c/n$  for some  $c = \text{poly}(1/\epsilon)$ .

**Proof** The proof is similar to that of Lemma 3.21. See Appendix G for details of the proof.

Now, we consider combinations of the cases considered so far and can show that the potential function,  $\Gamma$ , behaves as a super-martingale.

**Theorem 3.23** For the potential function,  $\Gamma$ ,  $\mathbb{E}[\Gamma(t+1)|x(t)] \leq (1 - \frac{\alpha.f^*}{16n(1+\epsilon\gamma_1)})\Gamma(t) + \frac{c}{n^2}$ , for constant  $c = \text{poly}(1/\epsilon)$ .

**Proof** We consider the following cases on intervals of values for  $s_i$ .

- **Case 1:**  $s_{(n\gamma_1)} \geq 0$  and  $s_{(n\gamma_2)} < 0$ . Using, Lemma 3.19 and Lemma 3.20, we can immediately see that,  $\mathbb{E}[\Gamma(t+1)|x(t)] \leq (1 - \alpha.f^*/8n)\Gamma(t)$  and hence, the result is also true.



- **Case 2:**  $s_{n\gamma_1} \geq s_{n\gamma_2} > 0$ . This represents a high load imbalance across the bins. In some cases,  $\Phi$  may grow but the asymmetry in the load implies that  $\Gamma$  is dominated by  $\Psi$ . Thus, the decrease in  $\Psi$  offsets the increase in  $\Phi$  and hence the expected change in  $\Gamma$  is negative. Specifically, if  $\mathbb{E}[\Delta\Phi|x] \leq \frac{-\alpha f^* \Phi}{8n}$ , then using Lemma 3.20 we get that  $\mathbb{E}[\Gamma(t+1)|x(t)] \leq (1 - \alpha f/8n)\Gamma(t)$ ; else we consider the following two cases:

- **Case 2a:**  $\Phi < \epsilon\gamma_1 * \Psi$ . Here, using Lemma 3.20 and Corollary 3.16, we get:

$$\begin{aligned}\mathbb{E}[\Delta\Gamma|x] &= \mathbb{E}[\Delta\Phi|x] + \mathbb{E}[\Delta\Psi|x] \\ &\leq \frac{\alpha.f^*}{n}.\Phi - \frac{\alpha.f^*}{2n} * \Psi \\ &\leq -\frac{\epsilon}{4n}.\Psi \\ &\leq -\frac{\epsilon}{4n(1 + \epsilon\gamma_1)}\Gamma\end{aligned}$$

- **Case 2b:**  $\Gamma < c/n$ . Here, using Corollary 3.18 and Corollary 3.16, we get:

$$\mathbb{E}[\Delta\Gamma|x] \leq \frac{\alpha.f^*}{n} * \Gamma \leq \frac{c\alpha.f^*}{n^2}$$

But,  $c/n^2 - ((\alpha.f^*/8n) * \Gamma) \geq c/n^2(1 - \alpha.f^*/8) \geq c/n^2(1 - \alpha.f^*/2) \geq \frac{c\alpha.f^*}{n^2}$ .  
Hence,  $\mathbb{E}[\Delta\Gamma|x] \leq -\frac{\alpha.f^*\Gamma}{8n} + \frac{c}{n}$ .

- **Case 3:**  $s_{n\gamma_2} \leq s_{n\gamma_1} < 0$ . Here, if  $\mathbb{E}[\Delta\Psi|x] \leq \frac{-\alpha.f^*}{2n}\Psi$ , then using Lemma 3.19, we get that  $\mathbb{E}[\Gamma(t+1)|x(t)] \leq (1 - \alpha.f^*/8n)\Gamma(t)$ ; else we consider the following two cases:

- **Case 3a:**  $\Psi < \epsilon\gamma_1 * \Phi$ . Here, using Lemma 3.19 and Corollary 3.18, we get:

$$\begin{aligned}\mathbb{E}[\Delta\Gamma|x] &= \mathbb{E}[\Delta\Phi|x] + \mathbb{E}[\Delta\Psi|x] \\ &\leq -(\alpha.f^*/8n)\Phi + \alpha.f^*/n * \Psi \\ &\leq -(\alpha.f^*/8n).\Phi + (\gamma_1\epsilon\alpha.f^*/n)\Phi \\ &\leq \frac{-\alpha.f^*}{16n}\Phi \\ &\leq \frac{-\alpha.f^*}{16n(1 + \epsilon\gamma_1)} * \Gamma\end{aligned}$$

- **Case 3b:**  $\Gamma < c/n$ . Here, using Corollary 3.16 and Corollary 3.18, we get:

$$\mathbb{E}[\Delta\Gamma|x] = \alpha.f^*/n * \Gamma \leq \frac{c\alpha.f^*}{n^2}$$

Hence, this case follows similarly as *Case 2b* above.

Now, we can prove using induction that the expected value of  $\Gamma$  remains bounded.

**Theorem 3.24** For any time  $t \geq 0$ ,  $\mathbb{E}[\Gamma(t)] \leq \frac{16c(1+\epsilon\gamma_1)}{n\alpha.f^*}$

**Proof** Using induction we can prove this claim. For  $t = 0$ , it is trivially true since  $\Gamma(0) \leq 2/n$ . Using Theorem 3.23, we get:

$$\begin{aligned}\mathbb{E}[\Gamma(t+1)] &= E[E[\Gamma(t+1)|\Gamma(t)]] \\ &\leq \mathbb{E}[(1 - \frac{\alpha.f^*}{16n(1 + \epsilon\gamma_1)})\Gamma(t) + \frac{c}{n}] \\ &\leq \frac{16c(1 + \epsilon\gamma_1)}{n\alpha.f^*} - \frac{c}{n^2} + \frac{c}{n^2} \\ &\leq \frac{16c(1 + \epsilon\gamma_1)}{n\alpha.f^*}\end{aligned}$$

**Theorem 3.25 Variable  $f$  Case (Weighted Case) Gap:** Using the bias in the probability distribution in favor of lightly loaded md-bins as obtained from the  $d$ -choice process, and assuming that in each ball, each dimension is chosen as 1 with probability  $q$  (variable  $f$  case); the expected and probabilistic upper bound on the gap (maximum dimensional gap) across the multidimensional bins is given as follows. Let,  $\delta = \frac{16c(1+\epsilon\gamma_1)}{\alpha \cdot f^*}$ , and  $\zeta > 0$ , then:

$$E[\text{Gap}(t)] \leq 2q \log(n)/\epsilon + 2q \log(\delta)/\epsilon$$

$$\Pr[\text{Gap}(t) > (mq/n)^{1/2+\zeta} (4q \log(n)/\epsilon + 4q \log(\delta)/\epsilon)] \leq \frac{1}{qm}$$

**Proof** Since, each dimension is assigned 1 with probability  $q$ , the average number of ones per md-ball is  $f^* = Dq$ . Let,  $a$  be the winning md-bin and  $m$  be the winning dimension that represents  $\text{Gap}(t)$ . The number of ones in any ball,  $f$ , follows a Binomial( $D, q$ ) distribution and has finite second moment. Using the analysis for the weighted balls case, we get,  $\mathbb{E}[e^{\alpha \cdot s_a}] \leq n\delta$ , where  $\alpha \cdot f^* \leq \epsilon/2$ . So,  $\mathbb{E}[e^{\alpha \cdot (x_a^m + \sum_{d \neq m} x_a^d)}] \leq n\delta$ . Taking, logarithm of both sides, we get:

$$E[x_a^m] + \sum_{d \neq m} E[x_a^d] \leq \log(n)/\alpha + \log(\delta)/\alpha \quad (3.31)$$

If  $k$  is the expected number of balls were thrown in bin  $a$  minus the average number of balls per bin, then  $E[x_a^m] = kq$  and similarly,  $\sum_{d \neq m} E[x_a^d] = (D-1)kq$ . Hence, we get:

$$Dkq \leq 2f^* \log(n)/\epsilon + 2f^* \log(\delta)/\epsilon$$

$$\Rightarrow k \leq 2 \log(n)/\epsilon + 2 \log(\delta)/\epsilon$$

$$\Rightarrow E[x_a^m] \leq 2q \log(n)/\epsilon + 2q \log(\delta)/\epsilon$$

The probabilistic bound can be computed similar to the fixed  $f$  case (Theorem 3.11) using the Chernoff bound.  $\square$

Note that the for the scalar case, when the expected weight of the distribution is  $W^*$ , the upper bound on the gap obtained is  $O(W^* \log(n))$ , which after normalization, i.e.  $E(W) = W^* = 1$ , leads to  $O(\log(n))$  gap. This improves upon the best prior known bound of  $O(n^c)$  given in [12].

## 4 $(1 + \beta)$ Choice Process with Multidimensional Balls and Bins

In this section we present upper and lower bounds on the gap for the  $(1 + \beta)$  choice process with multidimensional balls and bins.

### 4.1 Markov Chain Specification

As mentioned earlier, a balls-and-bins process can be characterized by a probability distribution vector  $(p_1, p_2, p_3, \dots, p_n)$ , where,  $p_i$  is the probability a ball is placed in the  $i^{\text{th}}$  most loaded multidimensional bin. Let  $x_i^d(t)$  be the random variable, that denotes the *weight in dimension  $d$  for bin  $i$*  and is equal to the load of the  $d^{\text{th}}$  dimension of the  $i^{\text{th}}$  bin minus the average load in dimension  $d$ . So,  $\sum_{i=1}^n x_i^d(t) = 0, \forall d \in [1..D]$ . Let,  $s_i(t)$  denote the sum of the loads (minus corresponding dimension averages) across all  $D$  dimensions for the bin  $i$  at time  $t$ , expressed as  $s_i(t) = \sum_{d=1}^D x_i^d(t)$ . It is assumed that bins are sorted by  $s_i(t)$ . So,  $s_i \geq s_{i+1} \forall i \in [1..n-1]$ . The process defines a Markov chain over the matrices,  $x(t)$  as follows:

- Sample  $j \in_p [n]$ .
- Set  $r_i = s_i(t) + f(1 - 1/n)$ , for  $i = j$ . Since, each md-ball has  $f$  non-zero entries, so each of these  $f$  dimensions in the bin,  $i$ , will be incremented by  $1 - 1/n$ .
- Set  $r_i = s_i(t) - f/n$ , for  $i \neq j$ . Since, each md-ball has  $f$  non-zero entries, so the each of the corresponding  $f$  dimensions in the bin,  $i$ , will be decremented by  $1/n$ . This ensures that for each dimension the sum across all the bins is 0.
- Obtain  $s(t+1)$  by sorting  $r(t)$ .

Fig. 1 (in the Appendix A) illustrates a multidimensional balls and bins scenario. The bounds on the gap will be proven for a family of probability distribution vectors  $p$ . As mentioned earlier, the md-bins are sorted based on their total dimensional load, i.e. sum of the weights across all dimensions for each bin ( $s_i$  for bin  $i$ ). We make the following assumptions:

- $\forall i \in [1, n-1], p_i \leq p_{i+1}$  This assumption states that the allocation rule is no worse than the 1-choice scheme.
- For some constants,  $\epsilon > 0, \theta > 1$  and  $0 < \gamma_3 < \gamma_4 < 1$ , where  $\gamma_3 + \gamma_4 = 1$ , it holds that:

$$p_{(n\gamma_3)} \leq \frac{(1 - \theta\epsilon)}{n}, \quad \text{and}, \quad p_{(n\gamma_4)} \geq \frac{(1 + \theta\epsilon)}{n} \quad (4.1)$$

This assumption states that the allocation rule strictly prefers the least loaded  $\gamma_3$  fraction of the  $n$  bins over the most loaded  $(1 - \gamma_4)$  fraction.

These assumptions imply that for some constants,  $\gamma_1$  and  $\gamma_2$ , where,  $0 < \gamma_1 < \gamma_3 < 1/2 < \gamma_4 < \gamma_2 < 1$  and  $\theta\gamma_1 = 1; \gamma_1 + \gamma_2 = 1$ , we have the following:

$\sum_{i \geq (n\gamma_2)} p_i \geq (\gamma_1 + \epsilon)$  and  $\sum_{i \leq (n\gamma_1)} p_i \leq (\gamma_1 - \epsilon)$ . This will be useful in the proof. Note that the  $(1 + \beta)$  choice process satisfies these assumptions for  $\epsilon = \beta(1 - 2\gamma_3)/\theta$ , since  $p_{(n\gamma_3)} \leq (1 - \beta)/n + 2(n\gamma_3 - 1)\beta/n^2 \leq (1 - \beta(1 - 2\gamma_3))/n$ , and similarly  $p_{(n\gamma_4)} \geq (1 + \beta(2\gamma_4 - 1))/n$ .

In the remaining analysis, we assume that when an md-ball arrives, then the selection of the bins is based on  $s_i$ , i.e. total sum of weights across all dimensions for the randomly selected bins (Fig. 1 in Appendix A). In particular, for the  $(1 + \beta)$  choice process, when two bins are randomly selected (with  $\beta$  probability), the md-ball (with  $f$  non-zero entries) is assigned to the md-bin with the lowest  $s_i$ . Using this selection mechanism, we prove the upper and lower bounds on the gap obtained for the  $(1 + \beta)$  choice process. Note that, this is a different allocation mechanism than that considered in [6] where the *max* objective is considered over the restricted set of  $f$  populated dimensions in the current md-ball.

## 4.2 Upper Bound On the Gap

We assume that  $\epsilon \leq 1/4$ . Further, let  $\alpha = \epsilon/2f$ . Define the following potential functions:

$$\begin{aligned} \Phi(t) &= \Phi(s(t)) = \sum_{i=1}^n e^{\alpha \cdot s_i} \\ \Psi(t) &= \Psi(s(t)) = \sum_{i=1}^n e^{-\alpha \cdot s_i} \\ \Gamma(t) &= \Gamma(s(t)) = \Phi(t) + \Psi(t) \end{aligned} \quad (4.2)$$

where,  $s_i(t) = \sum_{d=1}^D x_i^d(t)$

In the beginning, each dimension for each bin has 0 weight, thus  $s_i = 0, \forall i$  and hence,  $\Gamma(0) = 2n$ . We show that if  $\Gamma(x(t)) \geq an$  for some  $a > 0$ , then  $\mathbb{E}[\Gamma(t+1)|x(t)] \leq (1 - \frac{\epsilon^2(1-2\gamma_1)}{4n(1+\epsilon\gamma_1)}) * \Gamma(t)$ . This helps in demonstrating that for every given  $t$ ,  $\mathbb{E}[\Gamma(t)] \in O(n)$ . This implies that the maximum gap is  $O(\log(n))$  w.h.p.

First, consider the change in  $\Phi(t)$  (also refers to  $\Phi$  by default) and  $\Psi(t)$  (also refers to  $\Psi$  by default) separately when a ball is thrown with the given probability distribution.

**Lemma 4.1** *When an md-ball is thrown into an md-bin, the following inequality holds:*

$$\mathbb{E}[\Phi(t+1) - \Phi(t)|x(t)] \leq \sum_{i=1}^n [p_i * (\alpha \cdot f + (\alpha \cdot f)^2) - \alpha \cdot f/n] \cdot e^{\alpha \cdot s_i} \quad (4.3)$$

**Proof** Let  $\Delta_i$  be the expected change in  $\Phi$  if the ball is put in bin,  $i$ . So,  $r_i(t+1) = s_i + f(1 - 1/n)$ ; and for  $j \neq i$ ,  $r_j(t+1) = s_j(t) - f/n$ . The new values i.e.  $s(t+1)$  are obtained by sorting  $r(t+1)$  and  $\Phi(s) = \Phi(r)$ . The expected contribution of bin,  $i$ , to  $\Delta_i$  is given as follows:

$$\mathbb{E}[e^{\alpha \cdot (s_i + f(1-1/n))}] - e^{\alpha \cdot s_i} = e^{\alpha \cdot s_i} [e^{\alpha \cdot f(1-1/n)} - 1]$$

Similarly, the expected contribution of bin,  $j$  ( $j \neq i$ ) to  $\Delta_i$  is given as:

$$\mathbb{E}[e^{\alpha \cdot (s_j - f/n)}] - e^{\alpha \cdot s_j} = e^{\alpha \cdot s_j} [e^{-\alpha \cdot f/n} - 1]$$

Therefore,  $\Delta_i$  is given as follows:

$$\begin{aligned} \Delta_i &= e^{\alpha \cdot s_i} [e^{\alpha \cdot f(1-1/n)} - 1] + \sum_{j \neq i} e^{\alpha \cdot s_j} (e^{-\alpha \cdot f/n} - 1) \\ &= e^{\alpha \cdot (s_i - f/n)} (e^{\alpha \cdot f} - 1) + (e^{-\alpha \cdot f/n} - 1) \cdot \Phi \end{aligned}$$

Thus, we get the overall expected change in  $\Phi$  as follows:

$$\begin{aligned}
\mathbb{E}[\Phi(t+1) - \Phi(t)|x(t)] &= \sum_{i=1}^n p_i * \Delta_i \\
&= \sum_{i=1}^n p_i * [e^{\alpha(s_i - f/n)}(e^{\alpha f} - 1) + (e^{-\alpha f/n} - 1) \cdot \Phi] \\
&= \sum_{i=1}^n p_i * e^{\alpha(s_i - f/n)}(e^{\alpha f} - 1) + (e^{-\alpha f/n} - 1) \cdot p_i \cdot \Phi \\
&= \sum_{i=1}^n [p_i * e^{-\alpha f/n}(e^{\alpha f} - 1) + (e^{-\alpha f/n} - 1)] \cdot e^{\alpha s_i}
\end{aligned} \tag{4.4}$$

Now,  $e^{(-\alpha f/n)} * (e^{\alpha f} - 1)$  can be approximated as follows:

$$\begin{aligned}
e^{(-\alpha f/n)} \cdot (e^{\alpha f} - 1) &\leq (1 - \alpha f/n + (\alpha f/n)^2) * (1 + \alpha f + (\alpha f)^2 - 1) \\
&\sim \alpha f + (\alpha f)^2 + O((\alpha f)^2/n) \\
e^{(-\alpha f/n)} \cdot (e^{\alpha f} - 1) &\lesssim (\alpha f + (\alpha f)^2)
\end{aligned}$$

Above, since,  $(\alpha f)^2/n$  is very small for large  $n$ , we have ignored the small terms. Similarly,  $(e^{-\alpha f/n} - 1) \lesssim -\alpha f/n$ . Hence, the expected change in  $\Phi$  can be given by:

$$\mathbb{E}[\Phi(t+1) - \Phi(t)|x(t)] \leq \sum_{i=1}^n [p_i * (\alpha f + (\alpha f)^2) - \alpha f/n] \cdot e^{\alpha s_i} \tag{4.5}$$

Simplifying further and observing that  $\Phi_i$  decreases and  $p_i$  increases with increasing  $i$  from 1 to  $n$ , one gets the following Corollary.

**Corollary 4.2**  $\mathbb{E}[\Phi(t+1) - \Phi(t)|x(t)] \leq (\alpha f)^2 * \Phi/n$

**Proof** Since,  $p_i$  are increasing and  $\Phi_i$  are decreasing, the maximum value taken by RHS of equation (4.5) will be when  $p_i = 1/n$  for all  $i \in [1..n]$ . Simplifying, we get the result.

Similarly, the change in  $\Psi$  can be derived as follows.

**Lemma 4.3** *When an md-ball is thrown into an md-bin, the following inequality holds:*

$$\mathbb{E}[\Psi(t+1) - \Psi(t)|x(t)] \leq \sum_{i=1}^n [p_i * (-\alpha f + (\alpha f)^2) + \alpha f/n] \cdot e^{-\alpha s_i} \tag{4.6}$$

Further observing that  $p_i > 0$ , one gets the following Corollary.

**Corollary 4.4**  $\mathbb{E}[\Psi(t+1) - \Psi(t)|x(t)] \leq (\alpha f \cdot \Psi)/n$

In the next two lemmas, Lemma 4.5 and Lemma 4.6, we consider a reasonably balanced md-bins scenario. We show that for such cases, the expected potential decreases. Specifically, for  $s_{(n\gamma_2)} \leq 0$ , the expected value of  $\Phi$  decreases and for  $s_{(n\gamma_1)} \geq 0$ , the expected value of  $\Psi$  decreases.

**Lemma 4.5** *Let  $\Phi$  be defined as above. If  $s_{(n\gamma_2)}(t) \leq 0$  then,  $\mathbb{E}[\Phi(t+1)|x(t)] \leq (1 - \frac{\epsilon^2}{4n})\Phi + 1$*

**Proof** From equation (4.5), we get,

$$\begin{aligned}
\mathbb{E}[\Phi(t+1) - \Phi(t)|x(t)] &\leq \sum_{i=1}^n (p_i * (\alpha f + (\alpha f)^2) - \alpha f/n) \cdot e^{\alpha s_i} \\
&\leq \sum_{i < n\gamma_2} (p_i * (\alpha f + (\alpha f)^2) - \alpha f/n) \cdot e^{\alpha s_i} + \sum_{i \geq n\gamma_2} p_i * (\alpha f + (\alpha f)^2) \cdot e^0 \\
&\leq \sum_{i < n\gamma_2} (p_i * (\alpha f + (\alpha f)^2) - \alpha f/n) \cdot e^{\alpha s_i} + 1
\end{aligned} \tag{4.7}$$

The last inequality follows since  $\alpha.f < 1/2$  and  $\sum_{i \geq n\gamma_2} p_i < 1$ . Now, we need to upper bound the term  $\sum_{i < n\gamma_2} (p_i * (\alpha.f + (\alpha.f)^2).e^{\alpha.s_i})$ . Since  $p_i$  is non-decreasing and  $\Phi_i$  is non-increasing, the maximum value is achieved when  $\Phi_i = (\Phi/(n\gamma_2))$  for each  $i < n\gamma_2$ . Hence, the maximum value is:  $(\alpha.f + (\alpha.f)^2)(\gamma_2 - \epsilon)\Phi/(n\gamma_2)$ . Thus, the expected change in  $\Phi$  can be computed, using equation (4.7) and the above bound, as follows:

$$\begin{aligned}\mathbb{E}[\Phi(t+1) - \Phi(t)|x(t)] &\leq (\alpha.f + (\alpha.f)^2)(\gamma_2 - \epsilon)\Phi/(n\gamma_2) - \alpha.f/n * \Phi + 1 \\ &\leq (\alpha.f)^2.\Phi/n - (\alpha.f\epsilon\Phi/n\gamma_2) + 1 \\ &\leq \epsilon^2.\Phi/4n - \epsilon^2\Phi/(2n\gamma_2) + 1 \\ &\leq \frac{-\epsilon^2}{4n}\Phi + 1\end{aligned}\tag{4.8}$$

**Lemma 4.6** *Let  $\Psi$  be defined as above. If  $s_{(n\gamma_1)}(t) \geq 0$  then,  $\mathbb{E}[\Psi(t+1)|x(t)] \leq (1 - \frac{\epsilon^2}{4n})\Psi + 1$*

**Proof** The proof is similar to that of Lemma 4.5. See Appendix H for details of the proof.

Now, we consider the remaining cases and show that in case the load across the bins, at time  $t$ , is not reasonably balanced, then for  $s_{n\gamma_2} > 0$ , either  $\Psi$  dominates  $\Phi$  or the potential function is  $O(n)$ .

**Lemma 4.7** *Let,  $s_{(n\gamma_2)} > 0$  and  $\mathbb{E}[\Delta\Phi|x(t)] \geq -\epsilon^2\Phi/4n$ . Then, either  $\Phi < \epsilon\gamma_1 * \Psi$ , or  $\Gamma < cn$  for some  $c = \text{poly}(1/\epsilon)$ .*

**Proof** From equation (4.5), we get:

$$\begin{aligned}\mathbb{E}[\Delta\Phi|x(t)] &\leq \sum_{i=1}^n (p_i * (\alpha.f + (\alpha.f)^2) - \alpha.f/n).e^{\alpha.s_i} \\ &\leq \sum_{i \leq n\gamma_3} (p_i * (\alpha.f + (\alpha.f)^2) - \alpha.f/n).e^{\alpha.s_i} + \sum_{i > n\gamma_3} (p_i * (\alpha.f + (\alpha.f)^2) - \alpha.f/n) * e^{\alpha.s_i} \\ &\leq [((1 - \theta\epsilon)/n) * (\alpha.f + (\alpha.f)^2) - \alpha.f/n].\Phi_{(\leq n\gamma_3)} + (\alpha.f)^2.\Phi_{(> n\gamma_3)}/n \\ &\leq [-\epsilon^2/(2n\gamma_1) + \epsilon^2/4n].\Phi_{(\leq n\gamma_3)} + \epsilon^2.\Phi_{(> n\gamma_3)}/4n \\ &\leq [-\epsilon^2/(2n\gamma_1) + \epsilon^2/4n].\Phi + \epsilon^2\theta\Phi_{(> n\gamma_3)}/2n \quad \because \theta\gamma_1 = 1\end{aligned}\tag{4.9}$$

Now, since  $\mathbb{E}[\Delta\Phi|x(t)] \geq -\epsilon^2\Phi/4n$ , we get:  $\Phi \leq \Phi_{(> n\gamma_3)}/\gamma_2$ . Let,  $B = \sum_i \max(0, s_i)$ . Note,  $\sum_i s_i = 0$ , since for each dimension  $d$ , the update maintains that,  $\sum_{d=1}^D x_i^d(t) = 0$ . One can observe that,  $\Phi_{(> n\gamma_3)} \leq n\gamma_4 * e^{\alpha.B/(n\gamma_3)}$ , since  $\gamma_3 + \gamma_4 = 1$ . This implies that,  $\Phi \leq ((n\gamma_4)/(\gamma_2)) * e^{\alpha.B/(n\gamma_3)}$ .

Since,  $s_{(n\gamma_2)} > 0$ , so,  $\Psi \geq n\gamma_1 * e^{\alpha.B/(n\gamma_1)}$ . If  $\Phi < \epsilon\gamma_1 * \Psi$ , then we are done. Else,  $\Phi \geq \epsilon\gamma_1 * \Psi$ . This implies:

$$(n\gamma_4/\gamma_2) * e^{\alpha.B/(n\gamma_3)} \geq \Phi \geq \epsilon\gamma_1 * \Psi \geq \epsilon.n\gamma_1^2 * e^{\alpha.B/(n\gamma_1)}$$

Thus,  $e^{\alpha.B/n} \leq (\frac{\theta^2\gamma_4}{\epsilon\gamma_2})^{\frac{\gamma_1\gamma_3}{\gamma_3-\gamma_1}}$ . So,  $\Gamma \leq ((1 + \theta)/\epsilon) * \Phi \leq ((1 + \theta)/\epsilon) * (n\gamma_4/\gamma_2) * e^{\alpha.B/(n\gamma_3)} \leq ((1 + \theta)/\epsilon) * (n\gamma_4/\gamma_2) * (\frac{\theta^2\gamma_4}{\epsilon\gamma_2})^{\frac{\gamma_1}{\gamma_3-\gamma_1}}$ . Hence,  $\Gamma \leq cn$ , where,  $c = \text{poly}(1/\epsilon)$ .

In the Lemma below, we consider the case where the load across the bins at time,  $t$ , is not reasonably balanced, and  $s_{(n\gamma_1)} < 0$ . Here, we show that either  $\Phi$  dominates  $\Psi$  or the potential function is  $O(n)$ .

**Lemma 4.8** *Let,  $s_{(n\gamma_1)} < 0$  and  $\mathbb{E}[\Delta\Psi|x(t)] \geq -\epsilon^2\Phi/4n$ . Then, either  $\Psi < \epsilon\gamma_1 * \Phi$ , or  $\Gamma < cn$  for some  $c = \text{poly}(1/\epsilon)$ .*

**Proof** The proof is similar to that of Lemma 4.7. See Appendix I for details of the proof.

Now, we consider combinations of the cases considered so far and can show that the potential function,  $\Gamma$ , behaves as a super-martingale.

**Theorem 4.9** *For the potential function,  $\Gamma$ ,  $\mathbb{E}[\Gamma(t+1)|x(t)] \leq (1 - \frac{\epsilon^2(1-2\gamma_1)}{4n(1+\epsilon\gamma_1)})\Gamma(t) + c$ , for constant  $c = \text{poly}(1/\epsilon)$ .*

**Proof** We consider the following cases on intervals of values for  $s_i$ .

- **Case 1:**  $s_{n\gamma_1} \geq 0$  and  $s_{n\gamma_2} \leq 0$ . Using, Lemma 4.5 and Lemma 4.6, we can immediately see that,  $\mathbb{E}[\Gamma(t+1)|x(t)] \leq (1 - \epsilon^2/4n)\Gamma(t) + c$ , for constant  $c = \text{poly}(1/\epsilon)$  and hence, the result is also true.
- **Case 2:**  $s_{n\gamma_1} \geq s_{n\gamma_2} > 0$ . This represents a high load imbalance across the bins. In some cases,  $\Phi$  may grow but the asymmetry in the load implies that  $\Gamma$  is dominated by  $\Psi$ . Thus, the decrease in  $\Psi$  offsets the increase in  $\Phi$  and hence the expected change in  $\Gamma$  is negative. Specifically, if  $\mathbb{E}[\Delta\Phi|x] \leq -\epsilon^2/4n * \Phi$ , then using Lemma 4.6 we get that  $\mathbb{E}[\Gamma(t+1)|x(t)] \leq (1 - \epsilon^2/4n)\Gamma(t) + c$ ; else we consider the following two cases:

- **Case 2a:**  $\Phi < \epsilon\gamma_1 * \Psi$ . Here, using Lemma 4.6 and Corollary 4.2, we get:

$$\begin{aligned}\mathbb{E}[\Delta\Gamma|x] &= \mathbb{E}[\Delta\Phi|x] + \mathbb{E}[\Delta\Psi|x] \\ &\leq \frac{(\alpha.f)^2}{n} \cdot \Phi - \frac{\epsilon^2}{4n} * \Psi + 1 \\ &\leq -(1 - \epsilon\gamma_1) * \frac{\epsilon^2}{4n} \cdot \Psi + 1 \leq -\frac{\epsilon^2(1 - \epsilon\gamma_1)}{4n(1 + \epsilon\gamma_1)} \Gamma + 1\end{aligned}$$

- **Case 2b:**  $\Gamma < cn$ . Here, using Corollary 4.4 and Corollary 4.2, we get:

$$\mathbb{E}[\Delta\Gamma|x] \leq \alpha.f/n * \Gamma \leq c\alpha.f$$

But,  $c - ((\epsilon^2(1 - \epsilon\gamma_1)/4n) * \Gamma) \geq c(1 - \epsilon^2(1 - \epsilon\gamma_1)/4) \geq c(1 - \epsilon/2) \geq c\alpha.f$ . Hence, the result follows.

- **Case 3:**  $s_{n\gamma_2} \leq s_{n\gamma_1} < 0$ . Here, if  $\mathbb{E}[\Delta\Psi|x] \leq \frac{-\epsilon^2}{4n} \Psi$ , then using Lemma 4.5, we get that  $\mathbb{E}[\Gamma(t+1)|x(t)] \leq (1 - \epsilon^2/4n)\Gamma(t) + c$ ; else we consider the following two cases:
- **Case 3a:**  $\Psi < \epsilon\gamma_1 * \Phi$ . Here, using Lemma 4.5 and Corollary 4.4, we get:

$$\begin{aligned}\mathbb{E}[\Delta\Gamma|x] &= \mathbb{E}[\Delta\Phi|x] + \mathbb{E}[\Delta\Psi|x] \\ &\leq -(\epsilon^2/4n)\Phi + \alpha.f/n * \Psi + 1 \\ &\leq -(\epsilon^2/4n) \cdot \Phi + (\gamma_1\epsilon^2/2n)\Phi \\ &\leq \frac{\epsilon^2(\gamma_1 - 1/2)}{2n} \Phi + 1 \leq \frac{-\epsilon^2(1 - 2\gamma_1)}{(4n(1 + \epsilon\gamma_1))} * \Gamma + 1\end{aligned}$$

- **Case 3b:**  $\Gamma < cn$ . Here, using Corollary 4.2 and Corollary 4.4, we get:

$$\mathbb{E}[\Delta\Gamma|x] = \alpha.f/n * \Gamma \leq c\alpha.f$$

Hence, this case follows similarly as *Case 2b* above.

Now, we can prove using induction that the expected value of  $\Gamma$  remains bounded.

**Theorem 4.10** For any time  $t \geq 0$ ,  $\mathbb{E}[\Gamma(t)] \leq \frac{4c(1+\epsilon\gamma_1)}{\epsilon^2(1-2\gamma_1)}n$

**Proof** Using induction we can prove this claim. For  $t = 0$ , it is trivially true since  $\Gamma(0) = 2n$ . Using Theorem 4.9, we get:

$$\begin{aligned}\mathbb{E}[\Gamma(t+1)] &= E[E[\Gamma(t+1)|\Gamma(t)]] \\ &\leq \mathbb{E}[(1 - \frac{\epsilon^2(1 - 2\gamma_1)}{4n(1 + \epsilon\gamma_1)})\Gamma(t) + c] \\ &\leq \frac{4c(1 + \epsilon\gamma_1)}{\epsilon^2(1 - 2\gamma_1)}n - c + c\end{aligned}$$

Now, we can upper bound the gap across all the  $D$  dimensions across all the  $n$  md-bins. This gap is defined as follows:

$$Gap(t) = \max_{d=1}^D [\max_{i=1}^n x_i^d] \quad (4.10)$$

**Theorem 4.11 Fixed  $f$  Case:** Using the bias ( $p_{n\gamma_3} \leq (1 - \theta\epsilon)/n$  and  $p_{n\gamma_4} \geq (1 + \theta\epsilon)/n$ ) in the probability distribution in favor of lightly loaded md-bins, and assuming that  $f$  dimensions are exactly populated in each md-ball with uniform distribution of  $f$  dimensions over  $D$ , then the expected and probabilistic upper bound on the gap (maximum dimensional gap) across the multidimensional bins is given as follows. Let,  $\delta = \frac{4c(1+\epsilon\gamma_1)}{\epsilon^2(1-2\gamma_1)}$ , then:

$$\begin{aligned} E[\text{Gap}(t)] &\leq 2\log(n)/\epsilon + 2f\log(\delta)/\epsilon \\ \Pr[\text{Gap}(t) > 4\log(n)/\epsilon + 4\log(\delta)/\epsilon] &\leq 1/n \end{aligned}$$

**Proof** Let,  $a$  be the winning md-bin and  $m$  be the winning dimension that represents  $\text{Gap}(t)$ . Now, from Theorem 4.10, we get,  $\mathbb{E}[e^{\alpha \cdot s_a}] \leq n\delta$ . So,  $\mathbb{E}[e^{\alpha \cdot (x_a^m + \sum_{d \neq m} x_a^d)}] \leq n\delta$ . Let,  $y_a$  denote the gap as measured by the number of md-balls in bin  $a$  minus the average number of balls across the bins. Then,

$$\begin{aligned} E[s_a] &\leq 1/\alpha * \log(n) + 1/\alpha * \log(\delta) \\ \Rightarrow E[s_a] &\leq 2f\log(n)/\epsilon + O(2f\log(\delta)/\epsilon) \\ \Rightarrow f \cdot E[y_a] &\leq 2f\log(n)/\epsilon + O(2f\log(\delta)/\epsilon) \\ \Rightarrow E[y_a] &\leq 2\log(n)/\epsilon + O(2\log(\delta)/\epsilon) \end{aligned} \tag{4.11}$$

The third inequality uses the fact that each ball has exactly  $f$  populated dimensions. Since the  $f$  dimensions are chosen uniformly and randomly from  $D$  dimensions, the expected gap in any dimension (and hence the winning dimension with the maximum gap) is bounded by  $O(\frac{\log(n)}{\beta})$ , since  $\epsilon = \Theta(\beta)$  (section 4.1).

Now, the  $\Pr[s_a > 4f\log(n)/\epsilon + 2f/\epsilon * \log(\delta)] \leq \Pr[\Gamma(t) \geq nE[\Gamma(t)]] \leq 1/n$  (using Markov's Inequality); where  $s_a = \sum_{d=1}^D x_a^d$ . Hence,  $\Pr[y_a > 4\log(n)/\epsilon + 4\log(\delta)/\epsilon] \leq 1/n$ .

**Theorem 4.12 Variable  $f$  Case:** Using the bias ( $p_{n\gamma_3} \leq (1 - \theta\epsilon)/n$  and  $p_{n\gamma_4} \geq (1 + \theta\epsilon)/n$ ) in the probability distribution in favor of lightly loaded md-bins, and assuming that each dimension is chosen as 1 with probability  $q$  (non-fixed  $f$  case); the expected and probabilistic upper bound on the gap (maximum dimensional gap) across the multidimensional bins is given as follows. Let,  $\delta = \frac{4c(1+\epsilon\gamma_1)}{\epsilon^2(1-2\gamma_1)}$ , then:

$$\begin{aligned} E[\text{Gap}(t)] &\leq 2\log(n)/\epsilon + m/n(1 - q) + 2\log(\delta)/\epsilon \\ \Pr[\text{Gap}(t) > 4\log(n)/\epsilon + m/n(1 - q) + 4\log(\delta)/\epsilon] &\leq 1/n \end{aligned}$$

**Proof** Since, each dimension is assigned 1 with probability  $q$ , the average number of ones per md-ball is  $f^* = Dq$ . Let,  $a$  be the winning md-bin and  $m$  be the winning dimension that represents  $\text{Gap}(t)$ . The number of ones in any ball,  $f$ , follows a Binomial( $D, q$ ) distribution and has finite second moment. Using the analysis similar as for Theorem 4.10, we can get (proof omitted for brevity),  $\mathbb{E}[e^{\alpha \cdot s_a}] \leq n\delta$ , where  $\alpha \cdot f^* \leq \epsilon/2$ . So,  $\mathbb{E}[e^{\alpha \cdot (x_a^m + \sum_{d \neq m} x_a^d)}] \leq n\delta$ . Taking, logarithm of both sides, we get:

$$\begin{aligned} E[x_a^m] + \sum_{d \neq m} E[x_a^d] &\leq \log(n)/\alpha + \log(\delta)/\alpha \\ \Rightarrow E[l_a^m] + \sum_{d \neq m} E[l_a^d] - (D * \frac{mq}{n}) &\leq \log(n)/\alpha + \log(\delta)/\alpha \end{aligned} \tag{4.12}$$

In the second inequality,  $l_a^d$  represents the load in dimension  $d$  for bin  $a$ . Further, since the average load in each dimension is  $\frac{mq}{n}$ . If  $k$  balls were thrown in bin  $a$ , then  $E[l_a^m] = kq$  and similarly,  $\sum_{d \neq m} E[l_a^d] = (D - 1)kq$ . Hence, we get:

$$\begin{aligned} Dkq &\leq 2f^* \cdot \log(n)/\epsilon + 2f^* \log(\delta)/\epsilon + m f^* / n \\ \Rightarrow k &\leq 2\log(n)/\epsilon + 2\log(\delta)/\epsilon + m/n \\ \Rightarrow E[x_a^m] &\leq 2\log(n)/\epsilon + 2\log(\delta)/\epsilon + m/n(1 - q) \end{aligned}$$

The probabilistic bound can be computed similar to the fixed  $f$  case (Theorem 4.11).  $\square$

### 4.3 Lower Bound

We can show that the upper bound, for fixed  $f$  case with uniform distribution, proved in section 4.2 is tight to within  $f/D$  factor. Consider the case when,  $an\log(n)/\beta^2$  balls are thrown into  $n$  bins, using the  $(1 + \beta)$  choice process. The expected dimensional sum load per bin is  $af\log(n)/\beta^2$ . Now, the expected number of balls thrown using the

$(1 + \beta)$  choice process is  $an(1 - \beta) \log(n)/\beta^2$ . Raab and Steger [11] show that when  $cn \log(n)$  balls are thrown uniformly and randomly into  $n$  bins, then the load of the most loaded bin is at least  $(c + \sqrt{c}/10) \log(n)$  balls. Using,  $c = a(1 - \beta)/\beta^2$ , one can see that sum load in the maximum sum load bin is at least:

$$\left(\frac{a(1 - \beta)}{\beta^2} + \sqrt{\frac{a(1 - \beta)}{100\beta^2}}\right) * f \log(n) = \left(\frac{a}{\beta^2} + \frac{\sqrt{a(1 - \beta)} - a}{10\beta}\right) * f \log(n)$$

Since, each ball has  $f$  populated dimensions, hence, there are at least  $O(\log(n)/\beta + a \log(n)/\beta^2)$  balls in this max sum load bin. Since, in each ball  $f$  dimensions are uniformly distributed over  $D$  dimensions, there exists a dimension whose load is at least  $O(f \log(n)/D\beta)$  more than the average. Hence, the lower bound is  $O(f \log(n)/D\beta)$ .

## 5 Conclusions & Future Work

In this paper, we consider the challenging problem of multidimensional balanced allocation for both the sequential and the parallel  $d$  choice process and show that the gap (assuming fixed  $f$  populated dimensions per ball and uniform distribution of  $f$  over  $D$ ) is  $O(\log \log(n))$ , which is tight (within  $D/f$  factor of the lower bound). This improves the best prior [6] bound of  $O(\log \log(nD))$ . Further, for arbitrary number of balls  $m \gg n$ , the expected gap also has upper bound of  $O(\log \log(n))$ , that is independent of  $m$  for the fixed  $f$  case with uniform distribution of populated dimensions. For the variable  $f$  case with (non-uniform) binomial distribution of populated dimensions, the gap is  $O(\log(n))$  for  $m = O(n)$ . To the best of our knowledge, this is the first such analysis for  $d$ -choice paradigm with multidimensional balls and bins.

Our analysis also provides a much easier and elegant proof technique (as compared to [4]) for the  $O(\log \log(n))$  gap for  $m \gg n$  scalar balls thrown into  $n$  bins using the symmetric multiple choice process. Moreover, for the weighted sequential scalar balls and bins and general case  $m \gg n$ , we show the upper bound on the expected gap as  $O(\log(n))$  which improves upon the best prior bound of  $n^c$  ( $c$  depends on the weight distribution that has finite fourth moment) provided in [12]. In future, we would like to generalize the potential function approach for parallel and weighted balls and bins.

Further, we consider the challenging problem of multidimensional balanced allocation for the  $(1 + \beta)$  choice process and show that for arbitrarily large number of balls, the expected gap (assuming fixed  $f$  populated dimensions per ball and uniform distribution of  $f$  over  $D$ ) is  $O(\frac{\log(n)}{\beta})$ , which is tight (within  $D/f$  factor of the lower bound) and also independent of  $m$ . Further, the expected gap is also independent of  $m$  for non-uniform distribution of  $f$  dimensions over  $D$ , with fixed  $f$  per ball) and for random  $f$  with Binomial distribution.

## References

- [1] Micah Adler, Soumen Chakrabarti, Michael Mitzenmacher, and Lars Rasmussen. Parallel randomized load balancing. In *Proceedings of the 27th Annual ACM Symposium on Theory of Computing*, pages 238–247, 1995.
- [2] F. Meyer auf der Heide, C. Scheideler, and V. Stemmann. Exploiting storage redundancy to speed up randomized shared memory. *Theoretical Computer Science*, 162(2):245 – 281, 1996.
- [3] Yossi Azar, Andrei Z. Broder, Anna R. Karlin, and Eli Upfal. Balanced allocations. *SIAM Journal of Computing*, 29(1):180 – 200, 1999.
- [4] Petra Berenbrink, Artur Czumag, Angelika Steger, and Berthold Vocking. Balanced allocations: the heavily loaded case. *SIAM Journal of Computing*, 35(6):1350 – 1385, 2006.
- [5] Andrei Broder and Michael Mitzenmacher. Using multiple hash functions to improve ip lookups. Technical Report TR-03-00, Department of Computer Science, Harvard University, Cambridge, MA, 2000.
- [6] Andrei Broder and Michael Mitzenmacher. Multidimensional balanced allocations. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, SODA '05, pages 195–196, 2005.
- [7] R. Cole, B.M. Maggs, F. Meyer auf der Heide, M. Mitzenmacher, A.W. Richa, K. Schroder, R.K. Sitaraman, and B. Vocking. Randomized protocols for low congestion circuit routing in multi-stage interconnection networks. In *Thirteenth Annual Symposium ACM symposium on the Theory of Computing*, pages 378 – 388, May 1998.



- [8] Richard M. Karp, Michael Luby, and Friedhelm Meyer auf der heide. Efficient pram simulation on a distributed memory machine. In *24<sup>th</sup> Symposium on the Theory of Computing (STOC)*, pages 318 – 326, 1992.
- [9] Michael Mitzenmacher. *The Power of Two Choices in Randomized Load Balancing*. PhD in computer science, Harvard University, 1991.
- [10] Yuval Peres, Kunal Talwar, and Udi Wieder. The  $(1 + \beta)$ -choice process and weighted balls-into-bins. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '10, pages 1613–1619, 2010.
- [11] Martin Raab and Angelika Steger. Balls into bins - a simple and tight analysis. In *RANDOM*, pages 159 – 170, 1998.
- [12] Kunal Talwar and Udi Wieder. Balanced allocations: the weighted case. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, STOC '07, pages 256–265, New York, NY, USA, 2007. ACM.
- [13] B. Vocking. How asymmetry helps load balancing. In *FOCS*, pages 131 – 141, 1999.

## A Visualization of Multidimensional balls and bins

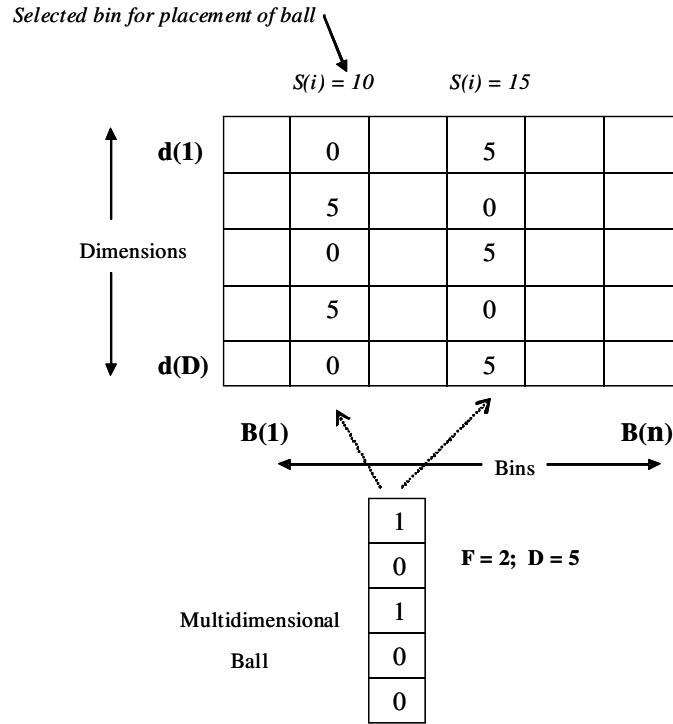


Figure 1: Multidimensional Balls and Bins Scenario

## B Proof of Lemma 3.3

The Lemma is restated below.

**Lemma B.1** When an  $md$ -ball is thrown into an  $md$ -bin, the following inequality holds:

$$\mathbb{E}[\Psi(t+1) - \Psi(t) | x(t)] \leq \sum_{i=1}^n [p_i * (-\alpha \cdot f + (\alpha \cdot f)^2) + \alpha \cdot f / n] \Psi_i \quad (\text{B.1})$$

**Proof** Let  $\Lambda_i$  be the expected change in  $\Psi$  if the ball is put in bin,  $i$ . So,  $r_i(t+1) = s_i + f(1 - 1/n)$ ; and for  $j \neq i$ ,  $r_j(t+1) = s_j(t) - f/n$ . The new values i.e.  $s(t+1)$  are obtained by sorting  $r(t+1)$  and  $\Psi(s) = \Psi(r)$ . The expected contribution of bin,  $i$ , to  $\Lambda_i$  is given as follows:

$$\mathbb{E}\left[\frac{e^{-\alpha \cdot (s_i + f(1-1/n))}}{p_0}\right] - \frac{e^{-\alpha \cdot s_i}}{p_0} = \frac{e^{-\alpha \cdot s_i}}{p_0} [e^{-\alpha \cdot f(1-1/n)} - 1]$$

Similarly, the expected contribution of bin,  $j$  ( $j \neq i$ ) to  $\Lambda_i$  is given as:

$$\mathbb{E}\left[\frac{e^{-\alpha \cdot (s_j - f/n)}}{j}\right] - \frac{e^{-\alpha \cdot s_j}}{j} = \frac{e^{-\alpha \cdot s_j}}{j} [e^{\alpha \cdot f/n} - 1]$$

Therefore,  $\Lambda_i$  is given as follows:

$$\begin{aligned} \Lambda_i &= \Psi_i [e^{-\alpha \cdot f(1-1/n)} - 1] + \sum_{j \neq i} \Psi_j (e^{\alpha \cdot f/n} - 1) \\ &= \Psi_i e^{\alpha f/n} (e^{-\alpha \cdot f} - 1) + (e^{\alpha \cdot f/n} - 1) \cdot \Psi \end{aligned}$$

Thus, we get the overall expected change in  $\Psi$  as follows:

$$\begin{aligned} \mathbb{E}[\Psi(t+1) - \Psi(t) | x(t)] &= \sum_{i=1}^n p_i * \Lambda_i \\ &= \sum_{i=1}^n p_i * [\Psi_i e^{\alpha f/n} (e^{-\alpha \cdot f} - 1) + (e^{-\alpha \cdot f/n} - 1) \cdot \Psi] \\ &= \sum_{i=1}^n p_i * \Psi_i e^{\alpha f/n} (e^{-\alpha \cdot f} - 1) + (e^{-\alpha \cdot f/n} - 1) \cdot p_i \cdot \Psi \\ &= \sum_{i=1}^n [p_i * e^{\alpha \cdot f/n} (e^{-\alpha \cdot f} - 1) + (e^{\alpha \cdot f/n} - 1)] \Psi_i \end{aligned} \tag{B.2}$$

Now,  $e^{\alpha \cdot f/n} (e^{-\alpha \cdot f} - 1)$  can be approximated as follows:

$$\begin{aligned} e^{\alpha \cdot f/n} \cdot (e^{-\alpha \cdot f} - 1) &\leq (1 + \alpha \cdot f/n + (\alpha \cdot f/n)^2) * (1 - \alpha \cdot f + (\alpha \cdot f)^2 - 1) \\ &\sim -\alpha \cdot f + (\alpha \cdot f)^2 + O((\alpha \cdot f)^2/n) \end{aligned}$$

Since,  $(\alpha \cdot f)^2/n$  is very small for large  $n$ , we ignore these small terms. Hence,

$$e^{\alpha \cdot f/n} \cdot (e^{-\alpha \cdot f} - 1) \lesssim (-\alpha \cdot f + (\alpha \cdot f)^2)$$

Similarly,  $(e^{\alpha \cdot f/n} - 1) \lesssim \alpha \cdot f/n$  Hence, the expected change in  $\Psi$  can be given by:

$$\mathbb{E}[\Psi(t+1) - \Psi(t) | x(t)] \leq \sum_{i=1}^n [p_i * (-\alpha \cdot f + (\alpha \cdot f)^2) + \alpha \cdot f/n] \Psi_i \tag{B.3}$$

## C Proof of Lemma 3.6

**Lemma C.1** Let  $\Psi$  be defined as above. If  $s_{(n\gamma_1)}(t) \geq 0$  then,  $\mathbb{E}[\Psi(t+1) | x(t)] \leq (1 - \frac{\alpha f}{8n}) \Psi$

**Proof** From equation (B.3), we get,

$$\begin{aligned} \mathbb{E}[\Psi(t+1) - \Psi(t) | x(t)] &\leq \sum_{i=1}^n (p_i * (-\alpha \cdot f + (\alpha \cdot f)^2) + \alpha \cdot f/n) \cdot \Psi_i \\ &\leq \sum_{i \geq n\gamma_1} p_i * (-\alpha \cdot f + (\alpha \cdot f)^2) \cdot \Psi_i + \alpha \cdot f/n * \Psi \end{aligned} \tag{C.1}$$

The last inequality follows since the  $(-\alpha \cdot f + (\alpha \cdot f)^2)$  is negative. Now, we need to upper bound the term  $\sum_{i \geq n\gamma_1} p_i * (-\alpha \cdot f + (\alpha \cdot f)^2) \cdot \Psi_i$ . Since,  $(-\alpha \cdot f + (\alpha \cdot f)^2)$  is negative, we need to find the minimum value of  $\sum_{i \geq n\gamma_1} p_i * \Psi_i$ .

Further,  $x_{n\gamma_1}(t) \geq 0$ , so  $\sum_{i \geq n\gamma_1} \Psi_i \geq \Psi - \ln(n\gamma_1)$ . Since  $p_i$  is non-decreasing and  $e^{-\alpha \cdot x_i}$  is non-decreasing, the minimum value of  $\sum_{i \geq n\gamma_1} p_i * \Psi_i$  is achieved, when,  $e^{-\alpha s_i} \ln(1/\gamma_1) = (\Psi - \ln(n\gamma_1))$  for each  $i \geq n\gamma_1$ . Thus, the minimum value is given as follows.

$$\begin{aligned} \sum_{i \geq n\gamma_1} p_i \Psi_i &\geq \frac{\Psi - \ln(n\gamma_1)}{\ln(1/\gamma_1)} \sum_{i \geq n\gamma_1} \frac{2i-1}{n^2} * \frac{1}{i} \\ &\geq \frac{\Psi - \ln(n\gamma_1)}{\ln(1/\gamma_1)} \left( \frac{2(1-\gamma_1)}{n} - \frac{\ln(1/\gamma_1)}{n^2} \right) \end{aligned} \quad (\text{C.2})$$

Thus, the expected change in  $\Psi$  can be computed, using equation (C.1) and the above bound, as follows:

$$\begin{aligned} \mathbb{E}[\Psi(t+1) - \Psi(t)|x(t)] &\leq (-\alpha \cdot f + (\alpha \cdot f)^2) * \frac{\Psi - \ln(n\gamma_1)}{\ln(1/\gamma_1)} \left( \frac{2(1-\gamma_1)}{n} - \frac{\ln(1/\gamma_1)}{n^2} \right) + \frac{\alpha \cdot f \Psi}{n} \\ &\leq \frac{-2\alpha f \Psi (1-\gamma_1)}{n \ln(1/\gamma_1)} - \frac{2\alpha \cdot f (1-\gamma_1) \ln(n\gamma_1)}{\ln(1/\gamma_1)} + \frac{\alpha f \Psi}{n} + O((\alpha \cdot f)^2) + O(1/n^2) \\ &\leq \frac{-\alpha f \Psi}{8n} \end{aligned} \quad (\text{C.3})$$

## D Proof of Lemma 3.8

**Proof** From equation (B.3), we get:

$$\begin{aligned} \mathbb{E}[\Delta \Psi | x(t)] &\leq \sum_{i=1}^n (p_i * (-\alpha \cdot f + (\alpha \cdot f)^2) + \alpha \cdot f/n) \cdot \Psi_i \\ &\leq \sum_{i > (n\gamma_4)} (p_i * (-\alpha \cdot f + (\alpha \cdot f)^2) + \alpha \cdot f/n) \cdot \Psi_i + \sum_{i \leq (n\gamma_4)} (p_i * (-\alpha \cdot f + (\alpha \cdot f)^2) + \alpha \cdot f/n) * \Psi_i \\ &\leq (-\alpha \cdot f + (\alpha \cdot f)^2) \cdot \frac{\Psi_{>(n\gamma_4)}}{\ln(1/\gamma_4)} * \sum_{i > (n\gamma_4)} \frac{2i-1}{n^2 i} + \\ &\quad (-\alpha \cdot f + (\alpha \cdot f)^2) \cdot \frac{\Psi_{\leq(n\gamma_4)}}{\ln(n\gamma_4)} * \sum_{i \leq (n\gamma_4)} \frac{2i-1}{n^2 i} + \frac{\alpha f \Psi}{n} \\ &\leq \frac{\alpha f \Psi (-2(1-\gamma_4) + \ln(1/\gamma_4))}{n \ln(1/\gamma_4)} + (2\alpha \cdot f \Psi_{\leq(n\gamma_4)}) \frac{-\gamma_4 \ln(1/\gamma_4) + (1-\gamma_4) \ln(n\gamma_4)}{n \ln(n\gamma_4) \ln(1/\gamma_4)} \end{aligned} \quad (\text{D.1})$$

In the above, the third inequality follows since,  $(-\alpha \cdot f + (\alpha \cdot f)^2)$  is negative and  $p_i \geq 0$ . Now, since  $\mathbb{E}[\Delta \Psi | x(t)] \geq \frac{-\alpha f \Psi}{8n}$ , we get that,  $\Psi \leq 4\Psi_{\leq(n\gamma_4)}$ . Let,  $B = \sum_i \max(0, s_i)$  (as mentioned in Lemma 3.7). One can observe that,  $\Psi_{\leq(n\gamma_4)} \leq \ln(n\gamma_4) * e^{\alpha \cdot B/(n\gamma_3)}$ . This implies that,  $\Psi \leq 4 \ln(n\gamma_4) e^{\alpha \cdot B/(n\gamma_3)}$ .

Since,  $s_{n\gamma_1} < 0$ , so,  $\Phi \geq \ln(n\gamma_1) * e^{\alpha \cdot B/(n\gamma_1)}$ . If  $\Psi < \epsilon \gamma_1 * \Phi$ , then we are done. Else,  $\Psi \geq \epsilon \gamma_1 * \Phi$ . This implies:

$$4 \ln(n\gamma_4) e^{\alpha \cdot B/(n\gamma_3)} \geq \Psi \geq \epsilon \gamma_1 * \Phi \geq (\epsilon \gamma_1 \ln(n\gamma_1)) * e^{\alpha \cdot B/(n\gamma_1)}$$

Thus,  $e^{\alpha \cdot B/n} \leq \left( \frac{4 \ln(n\gamma_4)}{\epsilon \gamma_1 \ln(n\gamma_1)} \right)^{\frac{\gamma_1 \gamma_3}{\gamma_3 - \gamma_1}}$ . So,  $\Gamma \leq ((\epsilon + \theta)/\epsilon) * \Psi \leq ((\epsilon + \theta)/\epsilon) * 4 \ln(n\gamma_4) e^{\alpha \cdot B/(n\gamma_3)}$ . Hence,  $\Gamma \leq c \ln(n)$ , where,  $c = \text{poly}(1/\epsilon)$ .

## E Proof of Lemma 3.17

The Lemma is restated below.

**Lemma E.1** When an  $md$ -ball is thrown into an  $md$ -bin, the following inequality holds:

$$\mathbb{E}[\Psi(t+1) - \Psi(t)|x(t)] \leq \sum_{i=1}^n [p_i * (-\alpha \cdot f^* + \frac{S\alpha^2}{n^2}) + \alpha \cdot f^*/n] \Psi_i \quad (\text{E.1})$$

**Proof** Let  $\Lambda_i$  be the expected change in  $\Psi$  if the ball is put in bin,  $i$ . So,  $r_i(t+1) = s_i + f(1 - 1/n)$ ; and for  $j \neq i$ ,  $r_j(t+1) = s_j(t) - f/n$ . The new values i.e.  $s(t+1)$  are obtained by sorting  $r(t+1)$  and  $\Psi(s) = \Psi(r)$ . When, an md-ball is committed to bin  $i$ , then it jumps to an index  $i_{new}$  which is less than or equal to  $i$  in the new bin order. Using similar analysis for taking care of these jumps as in Lemma 3.15, the expected contribution of bin,  $i$ , to  $\Lambda_i$  is given as follows:

$$\begin{aligned} \mathbb{E}\left[\frac{e^{-\alpha \cdot (s_i + f(1-1/n))}}{n^2 + n - i + 1}\right] - \frac{e^{-\alpha \cdot s_i}}{n^2 + n - i + 1} &= \frac{e^{-\alpha \cdot s_i}}{n^2 + n - i + 1} [M(-\alpha(1 - 1/n)) - 1] \\ &\leq \Psi_i(-\alpha \cdot f^*(1 - 1/n) + \frac{S\alpha^2}{n^2}) \end{aligned}$$

Similarly, the expected contribution of bin,  $j$  ( $j \neq i$ ) to  $\Lambda_i$  is given as:

$$\begin{aligned} \mathbb{E}\left[\frac{e^{-\alpha \cdot (s_j - f/n)}}{n^2 + n - j + 1}\right] - \frac{e^{-\alpha \cdot s_j}}{n^2 + n - j + 1} &= \frac{e^{-\alpha \cdot s_j}}{n^2 + n - j + 1} [M(\alpha/n) - 1] \\ &\leq \Psi_j \frac{-\alpha \cdot f^*}{n} \end{aligned}$$

Therefore,  $\Lambda_i$  is given as follows:

$$\Lambda_i = \Psi_i[\alpha \cdot f^* + \frac{S\alpha^2}{n^2}] + \frac{\alpha \cdot f^* \Psi}{n}$$

Hence, the expected change in  $\Psi$  can be given by:

$$\mathbb{E}[\Psi(t+1) - \Psi(t)|x(t)] \leq \sum_{i=1}^n [p_i * (-\alpha \cdot f^* + \frac{S\alpha^2}{n^2}) + \alpha \cdot f^*/n] \Psi_i \quad (\text{E.2})$$

## F Proof of Lemma 3.20

**Lemma F.1** Let  $\Psi$  be defined as above. If  $s_{(n\gamma_1)}(t) \geq 0$  then,  $\mathbb{E}[\Psi(t+1)|x(t)] \leq (1 - \frac{\alpha f^*}{2n})\Psi$

**Proof** From equation (E.1), we get,

$$\begin{aligned} \mathbb{E}[\Psi(t+1) - \Psi(t)|x(t)] &\leq \sum_{i=1}^n (p_i * (-\alpha \cdot f^* + \frac{S\alpha^2}{n^2}) + \alpha \cdot f^*/n) \cdot \Psi_i \\ &\leq \sum_{i \geq n\gamma_1} p_i * (-\alpha \cdot f^* + \frac{S\alpha^2}{n^2}) \cdot \Psi_i + \alpha \cdot f^*/n * \Psi \end{aligned} \quad (\text{F.1})$$

The last inequality follows since the  $(-\alpha \cdot f^* + \frac{S\alpha^2}{n^2})$  is negative. Now, we need to upper bound the term  $\sum_{i \geq n\gamma_1} p_i * (-\alpha \cdot f^* + \frac{S\alpha^2}{n^2}) \cdot \Psi_i$ . Since,  $(-\alpha \cdot f^* + \frac{S\alpha^2}{n^2})$  is negative, we need to find the minimum value of  $\sum_{i \geq n\gamma_1} p_i * \Psi_i$ . Since  $p_i$  is non-decreasing and  $\Psi_i$  is non-decreasing, the minimum value of  $\sum_{i \geq n\gamma_1} p_i * \Psi_i$  is achieved, when,  $e^{-\alpha s_i \frac{n-n\gamma_1}{n^2+1}} = \Psi$  for each  $i \geq n\gamma_1$ . Thus, the minimum value is given as follows.

$$\begin{aligned} \sum_{i \geq n\gamma_1} p_i \Psi_i &\geq \frac{n\Psi}{(1-\gamma_1)} \sum_{i \geq n\gamma_1} \frac{2i-1}{n^2} * \frac{1}{n^2 + n - i + 1} \\ &\geq \frac{\Psi}{n(1-\gamma_1)} * \frac{(2n+1)(1-\gamma_1)}{n} \end{aligned} \quad (\text{F.2})$$

Thus, the expected change in  $\Psi$  can be computed, using equation (F.1) and the above bound, as follows:

$$\begin{aligned} \mathbb{E}[\Psi(t+1) - \Psi(t)|x(t)] &\leq (-\alpha \cdot f^* + \frac{S\alpha^2}{n^2}) * \frac{\Psi}{n(1-\gamma_1)} * \frac{(2n+1)(1-\gamma_1)}{n} + \frac{\alpha \cdot f^* \Psi}{n} \\ &\leq \frac{-\alpha \cdot f^* \Psi}{n} + \frac{2S\alpha^2 \Psi}{n^2} - \frac{\alpha \cdot f^* \Psi}{n^2} \\ &\leq \frac{-\alpha \cdot f^* \Psi}{2n} \end{aligned} \quad (\text{F.3})$$

## G Proof of Lemma 3.22

The lemma is restated below:

**Lemma G.1** *Let,  $s_{(n\gamma_1)} < 0$  and  $\mathbb{E}[\Delta\Psi|x(t)] \geq -\alpha.f^*\Psi/2n$ . Then, either  $\Psi < \epsilon\gamma_1 * \Phi$ , or  $\Gamma < c/n$  for some  $c = \text{poly}(1/\epsilon)$ .*

**Proof** From equation (E.1), we get:

$$\begin{aligned}
\mathbb{E}[\Delta\Psi|x(t)] &\leq \sum_{i=1}^n (p_i * (-\alpha.f^* + \frac{S\alpha^2}{n^2} + \frac{\alpha.f^*}{n}).\Psi_i \\
&\leq \sum_{i>(n\gamma_4)} (p_i * (-\alpha.f^* + \frac{S\alpha^2}{n^2}) + \alpha.f^*/n).\Psi_i + \sum_{i\leq(n\gamma_4)} (p_i * (-\alpha.f^* + \frac{S\alpha^2}{n^2}) + \alpha.f^*/n) * \Psi_i \\
&\leq \frac{\alpha.f^*\Psi_{(\geq n\gamma_4)}}{n\gamma_4} * \frac{(2n+1)(\gamma_4-1)}{n} + \frac{\alpha.f^*}{n} \\
&\quad \frac{\alpha.f^*\Psi_{(< n\gamma_4)}}{n(\gamma_4-1)} * \frac{(2n+1)\gamma_4}{n} \\
&\leq \frac{\alpha.f^*\Psi}{n\gamma_4} * \frac{(2n+1)(\gamma_4-1)}{n} + \frac{\alpha.f^*}{n} \\
&\quad \frac{2\alpha.f^*\Psi_{(< n\gamma_4)}}{n\gamma_4(\gamma_4-1)} * \frac{(1-2\gamma_4)}{(1-\gamma_4)}
\end{aligned} \tag{G.1}$$

In the above, the third inequality follows since,  $(-\alpha.f^* + \frac{S\alpha^2}{n^2})$  is negative and  $p_i \geq 0$ . Now, since  $\mathbb{E}[\Delta\Psi|x(t)] \geq \frac{-\alpha.f^*\Psi}{2n}$ , we get:  $\Psi \leq \Psi_{(\geq n\gamma_4)} * \frac{4(2\gamma_4-1)}{\gamma_4(1-\gamma_4)(4\gamma_4-1)}$ . Let,  $B = \sum_i \max(0, s_i)$  (as mentioned in Lemma 3.21). One can observe that,  $\Psi_{(\geq n\gamma_4)} \leq * \frac{1-\gamma_4}{n} e^{\alpha.B/(n-n\gamma_4)}$ . This implies that,  $\Psi \leq \frac{4(2\gamma_4-1)}{\gamma_4(4\gamma_4-1)} e^{\alpha.B/(n-n\gamma_4)}$ .

Since,  $s_{n\gamma_1} < 0$ , so,  $\Phi \geq \frac{\gamma_1}{n} * e^{\alpha.B/(n\gamma_1)}$ . If  $\Psi < \epsilon\gamma_1 * \Phi$ , then we are done. Else,  $\Psi \geq \epsilon\gamma_1 * \Phi$ . This implies:

$$\frac{4(2\gamma_4-1)}{\gamma_4(4\gamma_4-1)} e^{\alpha.B/(n-n\gamma_4)} \geq \Psi \geq \epsilon\gamma_1 * \Phi \geq \frac{\epsilon\gamma_1^2}{n} * e^{\alpha.B/(n\gamma_1)}$$

Thus,  $e^{\alpha.B/n} \leq (\frac{4(2\gamma_4-1)}{\epsilon\gamma_1^2\gamma_4(4\gamma_4-1)})^{\frac{\gamma_1(1-\gamma_4)}{1-\gamma_4-\gamma_1}}$ . So,  $\Gamma \leq ((1+\theta)/\epsilon) * \Psi \leq ((1+\theta)/\epsilon) * \frac{4(2\gamma_4-1)}{n\gamma_4(4\gamma_4-1)} e^{\alpha.B/(n\gamma_3)}$ . Hence,  $\Gamma \leq \frac{c}{n}$ , where,  $c = \text{poly}(1/\epsilon)$ .

## H Proof of Lemma 4.6

**Lemma H.1** *Let  $\Psi$  be defined as above. If  $s_{(n\gamma_1)}(t) \geq 0$  then,  $\mathbb{E}[\Psi(t+1)|x(t)] \leq (1 - \frac{\epsilon^2}{4n})\Psi + 1$*

**Proof** From equation (4.6), we get,

$$\begin{aligned}
\mathbb{E}[\Psi(t+1) - \Psi(t)|x(t)] &\leq \sum_{i=1}^n (p_i * (-\alpha.f + (\alpha.f)^2) + \alpha.f/n).e^{-\alpha.s_i} \\
&\leq \sum_{i \geq n\gamma_1} p_i * (-\alpha.f + (\alpha.f)^2).e^{-\alpha.s_i} + \alpha.f/n * \Psi
\end{aligned} \tag{H.1}$$

The last inequality follows since the  $(-\alpha.f + (\alpha.f)^2)$  is negative. Now, we need to upper bound the term  $\sum_{i \geq n\gamma_1} p_i * (-\alpha.f + (\alpha.f)^2).e^{-\alpha.s_i}$ . Since,  $(-\alpha.f + (\alpha.f)^2)$  is negative, we need to find the minimum value of  $\sum_{i \geq n\gamma_1} p_i * e^{-\alpha.s_i}$ . Further,  $x_{n\gamma_1}(t) \geq 0$ , so  $\sum_{i \geq n\gamma_1} \Psi_i \geq \Psi - n\gamma_1$ . Since  $p_i$  is non-decreasing and  $\Psi_i$  is non-decreasing, the minimum value of  $\sum_{i \geq n\gamma_1} p_i * e^{-\alpha.s_i}$  is achieved, when,  $\Psi_i = (\Psi - n\gamma_1)/n\gamma_2$  for each  $i \geq n\gamma_1$ . Using the assumption, that  $\sum_{i \geq n\gamma_1} p_i \geq (\gamma_2 + \epsilon)$ , the minimum value is:  $(\gamma_2 + \epsilon)(\Psi - n\gamma_1)/(n\gamma_2)$ . Thus,

$$\sum_{i \geq n\gamma_1} p_i * (-\alpha.f + (\alpha.f)^2).e^{-\alpha.s_i} \leq (-\alpha.f + (\alpha.f)^2) * (\gamma_2 + \epsilon) * (\Psi - n\gamma_1)/(n\gamma_2) \tag{H.2}$$

Thus, the expected change in  $\Psi$  can be computed, using equation (H.1) and the above bound, as follows:

$$\begin{aligned}
\mathbb{E}[\Psi(t+1) - \Psi(t)|x(t)] &\leq (-\alpha.f + (\alpha.f)^2)(\gamma_2 + \epsilon)(\Psi - n\gamma_1)/(n\gamma_2) + \alpha.f/n * \Psi \\
&\leq \frac{-\epsilon^2}{2n\gamma_2}\Psi + \frac{\epsilon^2}{4n}\Psi + 1 \\
&\leq \frac{-\epsilon^2}{4n}\Psi + 1
\end{aligned} \tag{H.3}$$

## I Proof of Lemma 4.8

**Proof** From equation (4.6), we get:

$$\begin{aligned}
\mathbb{E}[\Delta\Psi|x(t)] &\leq \sum_{i=1}^n (p_i * (-\alpha.f + (\alpha.f)^2) + \alpha.f/n) \cdot e^{-\alpha.s_i} \\
&\leq \sum_{i > (n\gamma_4)} (p_i * (-\alpha.f + (\alpha.f)^2) + \alpha.f/n) \cdot e^{-\alpha.s_i} + \sum_{i \leq (n\gamma_4)} (p_i * (-\alpha.f + (\alpha.f)^2) + \alpha.f/n) * e^{-\alpha.s_i} \\
&\leq [(1 + \theta\epsilon)/n * (-\alpha.f + (\alpha.f)^2) + \alpha.f/n] \cdot \Psi_{>(n\gamma_4)} + (\alpha.f) \cdot \Psi_{\leq n\gamma_4}/n \\
&\leq [-\alpha f \theta \epsilon^2/n + \epsilon^2/4n] \cdot \Psi_{>(n\gamma_4)} + \alpha.f \cdot \Psi_{\leq(n\gamma_4)}/n \\
&\leq (-\epsilon^2/(2n\gamma_1) + \epsilon^2/(4n)) * \Psi + (\alpha.f/n + \epsilon^2/(2n\gamma_1)) \cdot \Psi_{\leq(n\gamma_4)}
\end{aligned} \tag{I.1}$$

In the above, the third inequality follows since,  $(-\alpha.f + (\alpha.f)^2)$  is negative and  $p_i \geq 0$ ; and  $\forall i > (n\gamma_4), p_i > (1 + \theta\epsilon)/n$ . Now, since  $\mathbb{E}[\Delta\Psi|x(t)] \geq -\epsilon^2\Psi/4n$ , we get that,  $\epsilon^2\gamma_2\Psi/(2n\gamma_1) \leq (\epsilon/2n + \epsilon^2/(2n\gamma_1)) * \Psi_{\leq(n\gamma_4)}$ . Thus, we get:  $\Psi \leq \frac{(1+\theta\epsilon)}{\epsilon\gamma_2} * \Psi_{\leq(n\gamma_4)}$ . Let,  $B = \sum_i \max(0, s_i)$  (as mentioned in Lemma 4.7). One can observe that,  $\Psi_{\leq(n\gamma_4)} \leq (n\gamma_4) * e^{\alpha.B/(n\gamma_3)}$ . This implies that,  $\Psi \leq \frac{(1+\theta\epsilon)}{\epsilon\gamma_2} * (n\gamma_4) * e^{\alpha.B/(n\gamma_3)}$ .

Since,  $s_{n\gamma_1} < 0$ , so,  $\Phi \geq n\gamma_1 * e^{\alpha.B/(n\gamma_1)}$ . If  $\Psi < \epsilon\gamma_1 * \Phi$ , then we are done. Else,  $\Psi \geq \epsilon\gamma_1 * \Phi$ . This implies:

$$((n(1 + \theta\epsilon)\gamma_4)/(\epsilon\gamma_2)) \cdot e^{\alpha.B/(n\gamma_3)} \geq \Psi \geq \epsilon\gamma_1 * \Phi \geq (\epsilon\gamma_1^2 n) * e^{\alpha.B/(n\gamma_1)}$$

Thus,  $e^{\alpha.B/n} \leq \left(\frac{(1+\theta\epsilon)\gamma_4}{\epsilon^2\gamma_2\gamma_1^2}\right)^{\frac{\gamma_1\gamma_3}{\gamma_3-\gamma_1}}$ . So,  $\Gamma \leq ((1 + \theta)/\epsilon) * \Psi \leq ((1 + \theta)/\epsilon) * \frac{(1+\theta\epsilon)\gamma_4 n}{\epsilon\gamma_2} \cdot e^{\alpha.B/(n\gamma_3)}$ . Hence,  $\Gamma \leq cn$ , where,  $c = \text{poly}(1/\epsilon)$ .